

Ethical watchdog's just desserts

The US scientific community continues to make heavy weather of the policing of scientific misconduct, chiefly because federal agencies cannot act without being forced to be courts of law. But now there is a chance to undo one mistake.

ANY organization called an Office of Scientific Integrity (OSI) is dangerously named. The US version of this institution (there could hardly be one elsewhere) is also named euphemistically; its creation was the response of the National Institutes of Health (NIH) to complaints from the Congress about what then seemed an endless rash of misconduct scandals. If NIH had called its new offspring, which is constitutionally autonomous, the Office of Scientific Misconduct, it might have avoided the trouble that now unexpectedly afflicts it — the decision of a Wisconsin trial judge that OSI's procedures for inquiring into cases of misconduct are beyond its legal powers (see page 357). Could the "positive thinking" embodied in OSI's title have helped to blind it to the judicial character of its own function?

That allegations of misconduct in the treatment and publication of data and their interpretation need to be investigated goes without saying. Too many institutions have found to their embarrassment in the past decade that scandals swept under the rug tend to be magnified in the process. But it has never been self-evident that NIH centrally is best placed to investigate such cases. For one thing, it is necessarily remote. For another, it is necessarily bureaucratic — which means its machinery moves slowly.

On the face of things, Judge Barbara Crabb has now thrown a spanner in these works by accepting a plaintiff's claim that OSI has behaved as if it were exempt from the Administrative Procedures Act, which applies to all US federal agencies (see *Nature* 349, 95; 10 January 1991). Those concerned are now wondering whether to legitimize what OSI does by advertising its procedures in the *Federal Register*, by allowing time for public comment and by putting up with all the inconvenience and delay that would result. It would be better, even at this late stage, that the NIH should ask themselves the questions they should have faced more courageously at the outset — is OSI (however named) the best way of calling a halt to scandalous behaviour in US biomedical research?

The NIH seem a natural choice for such a role for two reasons.

First, they have a hand in supporting most biomedical research in the United States (research with fetal tissue is a disgraceful exception) while the contractual relationship between a federal grant-maker and its recipients gives the former the right, even the duty, to ensure that its

funds are in every way properly spent. ('Accountability' is the buzz-word.) But NIH are also the lightning-rod for Congressional discontent. Asking for more funds each year would be that much more difficult without a seemingly display of zeal in attending to congressional anxieties.

Yet neither NIH nor their autonomous offspring OSI are equipped to be a policeman of biomedical research. OSI's well-meant efforts to simplify its procedures by making them more flexible and less formal than those of the courts have inevitably laid it open to the charge of denying due process to those against whom allegations of misconduct are made. But they are so varied in character and seriousness that they are best dealt with nearer to the laboratory bench, ideally by the university or research institution directly concerned. For the home institution is also best placed to act on the outcome of misconduct inquiries. At one extreme, it may be necessary that employees should be dismissed, but it is important that even such decisions should be tempered by a knowledge of extenuating circumstances, which home institutions are best placed to understand.

But NIH could and should use their contractual influence to ensure that institutions faced with allegations against their own researchers respond appropriately. That means acting quickly, appointing independent panels of inquiry except in the most trivial cases and giving a full report to NIH, which might then at its discretion make it public. There would still be a role for something like OSI to inquire into allegations against NIH employees, but by cherishing the ambition that OSI might act nationally and, by extension, globally, the NIH have hostages to unkind fortune. Judge Crabb may have given them a chance to think again. □

Overheads overblown

British researchers are being saddled with a system of research accounting that is failing in the United States.

UNIVERSITIES everywhere compete for research grants and often put out joyful press announcements when they succeed, but then are sometimes heard complaining of the cost of giving houseroom to the successful projects. That, at least, is the case in the United States, where both public and private universities are financed to teach their stu-



dents, not to undertake research. Yet everybody knows that the cost of mounting research projects is not small. Research grants do not flow readily to poorly equipped laboratories, people must spend time writing grant applications, the recruitment of technical staff (even when salaries are covered by grants) may be substantial and there is also heat and light to be paid for. In other words, there is an overhead that somebody must pay for. The US tradition is that, with greater or lesser grace, the grant-making agencies eventually stump up. But now, at the instance of the federal government, they are jibbing (see page 361). What lessons should British universities, about to embark on the same path (see page 339), learn from the US experience?

One is that, despite general agreement about the propriety of the research overhead (curiously called an "indirect cost" in the United States), there is general disagreement about its proper size. Second, there can be no single percentage charge that will equitably cover all universities; the running costs of city-centre campuses may be unusually high, for example, but they may be the better able to recruit researchers on that account. Third, telling what charge is equitable cannot be determined by measurement, which would entail (among other things) knowing how academics spend their time. Fourth, the end-result must be a matter for negotiation, even bargaining, between the finance officers of grant-making agencies and university administrators (who will be tempted to charge the time they spend against the grants won by their academic colleagues).

It is too soon to tell whether the noises coming from the US Congress this past budget cycle mark a change in the rules or are simply a novel element in a familiar ritual, but the research enterprise as a whole has more to gain than to lose from a clearer understanding with the federal agencies. One danger, for example, is that ambitious university administrators may overspend on laboratory buildings in the expectation that their investments will eventually be returned as overhead charges, forgetting that in the process they deprive the system of funds that could be used for research itself, or for helping make good the general shortage of trained people. Seemly rules would help everybody, restrictive though they might seem.

The British dilemma is, as usual, at the other end of the spectrum. British universities are only just abandoning the convention that all universities are equal, and equally capable of carrying out excellent research. This development has been foisted on the system by two developments — the general pressure on recurrent budgets over the past decade, which has forced many institutions to spend on routine teaching and departmental functions funds that should have been spent on the infrastructure of research, and then, in the past five years, self-conscious speculation fostered by government agencies about the distinctions between "research" and "teaching" universities. It is hardly surprising that a canny government has seized the opportunity to say that some of the university

budget should be transferred to the research councils, which should they pay overhead costs on some basis yet to be determined. The government will be able to boast of greater accountability, universities stroing in research may be marginally better off, but in British circumstances the general benefit is at best obscure.

Simple arithmetic shows the snags with the proposed arrangements.

For the past decade, it has been supposed that 30 per cent of recurrent university spending was devoted to the infrastructure of research, but that figure derives from the report by the late Sir Alec Merrison published in 1982 for which, even at the time, there was hardly any evidence. As some universities have devoted increasing proportions of their energy to teaching (there is nothing wrong with that, for that is their prime function), spending on the infrastructure of research has inevitably declined. And it shows. The result is that public support for academically based research has probably been declining even more quickly than can be told from the direct cost of research grants.

Yet curiously the government still counts 30 per cent of the cost of running universities in its annual summaries of public support for British science. British universities might usefully pick over this potential quarrel before they settle for what may otherwise be a meagre ration of overhead. □

Is this the eco-war?

Spilled oil in the Persian Gulf is better than spilled blood on the sand, but there is probably that ahead.

If the oil-spill in the Persian Gulf last week is said to be the worst ever, nobody should dissent. The president of Iraq has said from the beginning that foreign nationals now in Saudi Arabia would have to walk through fire if they were serious about reoccupying Kuwait; if United States aircraft had not set the off-shore terminal on fire by accident last week, there is every chance that fire would have been started deliberately instead. But even crude oil is flammable, so that a natural accident might well have done the trick. Wars have a tendency to evoke the unexpected; things go wrong.

If the oil-spill is described as a catastrophe, on the other hand, assent should be withheld, at least for the time being. It is true that some unique species are threatened by the southwards-spreading oil-slick, while it is sad to see (on television) so many cormorants at death's door. But we shall be lucky if the long-term consequences of the oil-slick, for surface or bottom species, outlast the human problems that there will be when the present conflict comes to an end. It is remarkable that the people who are energetically placing floating booms to protect desalination plants along the east-facing coastline of Saudi Arabia should be giving so little thought to the persistent problems of the Middle East. □

UK nuclear physicists fear SERC's cuts

- Nuclear-structure research threatened
- European collaborations in jeopardy

Daresbury, Cheshire

NUCLEAR structure research in Britain could be the latest victim of the struggle of the Science and Engineering Research Council (SERC) to cut £40 million from its planned spending in 1991–92, a crisis meeting of the British nuclear physics community heard last week.

Inflation means that SERC's share of this year's science budget (see page 359) will be worth less than in 1990–91 — whence SERC's strenuous effort to cut its research programme. But the more than 200 nuclear-structure physicists (some 80 per cent of the British community) gathered angrily at SERC's Daresbury Laboratory last week believe their field is being unfairly targeted for cuts by a small policy group, led by SERC chairman Sir Mark Richmond.

The five-member group, which includes no physicists, will put cost-cutting options to the full council on 6 February, when SERC must decide which parts of its research programme to cut if the Treasury fails to provide extra for next year.

Peter Twin, physics professor at the University of Liverpool and chairman of SERC's Nuclear Structure Committee, said last week that the policy group is likely to recommend a cut of nearly 10 per cent in the 1991–92 budget of SERC's Nuclear Physics Board, which channels funds to nuclear and high-energy physics.

Although similar cuts are expected across most of SERC, the problem for the Nuclear Physics Board is that almost two-thirds of its £90 million annual spending is the British subscription to CERN, the European high-energy physics laboratory at Geneva. The Foreign Office is said to have decreed that the CERN contribution cannot be renegotiated, for fear of souring Britain's relations with Europe.

Twin's fear is that the intended £8 million cut in Nuclear Physics Board spending will fall mainly on nuclear structure research. Of the board's unencumbered £30 million, almost £20 million supports particle physics, largely at CERN, where full support for British users is essential to justify continued membership. The "obvious" conclusion, said Twin, is that the cuts will be mostly targeted at the £10.1 million annual spending on nuclear-structure research, "completely demolishing our budget".

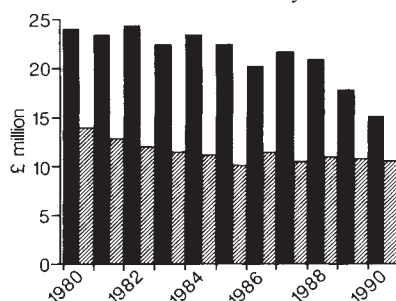
The first result of the threatened cut would be the closure of the tandem accelerator at the Daresbury Laboratory, opened in 1983, but still offering five or

more years useful service. Last week's gathering insisted that, without this machine, which revived British nuclear structure research after Britain missed out on a generation of accelerators in the 1970s, the future is bleak.

The nuclear-structure researchers' fears will be fuelled by a Nuclear Physics Board planning document, dated 25 January and obtained by *Nature*, which shows the £5 million annual running cost of the Daresbury tandem discontinued after 1991–92.

Nuclear-structure physicists are also outraged that the effective closure of a successful branch of British science could be contemplated on purely financial grounds, without scientific review. Last week they claimed a solid international reputation won in the 1980s for research in the structure of rare nuclei, despite declining support (see figure). Tony Hughes, responsible for SERC's own laboratories, said that other areas of spending are also seriously threatened, but conceded that the effective withdrawal of support from a particular field is unprecedented.

Future European projects are also at risk. SERC has agreed to collaborate with the French Centre National de la Recherche Scientifique (CNRS) in EURO-GAM, a £5-million project to build the world's most sensitive γ -ray detector array (*Nature* 344, 94; 1990). The plan is to build the device at Daresbury and to run



SERC's Nuclear Physics Board expenditure in the eighties on particle physics (solid bars) and nuclear structure (cross-hatched).

some initial experiments on the tandem accelerator before moving EURO-GAM to the CNRS Institute for Nuclear Research at Strasbourg.

Bernard Haas from Strasbourg said in Daresbury last week that the first phase of this project would continue, but the threatened budget cut would make the second Strasbourg phase "totally impossible". Europe would then lose its lead in the field to the United States, where a similar device is being planned, he said.

Haas also said that "the British would

SCIENTIFIC MISCONDUCT

Court ruling put to test

Washington

A Wisconsin scientist will be the first to test last month's court decision that the misconduct regulations of the National Institutes of Health (NIH) are illegal.

L. Cass Terry, a Medical College of Wisconsin neurologist being investigated by NIH for alleged scientific misconduct filed a lawsuit earlier this month, claiming that the inquiry (now in its fourth year) violates his right to fair treatment. Now, citing last month's decision in the case of James Abbs (see *Nature* 349, 95; 10 January 1991), Terry's lawyers plan to file a motion this week to have his investigation halted entirely.

The Abbs decision held that the NIH rules are invalid because they were not issued with an opportunity for public comment. NIH has not yet decided how to respond to the decision, and has neither appealed against it nor halted current investigations. But as similar law-suits based on the Abbs ruling emerge, it would seem that NIH cannot ignore the challenge.

Yet Suzanne Hadley, deputy director of the NIH Office of Scientific Integrity, says, "We do not intend to offer anybody any [special treatment] at this time". But because of the ruling, officials at NIH's parent agency, the Department of Health and Human Services, say they are preparing to recommend that NIH "do the right thing" and propose new regulations for public comment, a process that threatens to take years and further confuse the current state of misconduct affairs.

Christopher Anderson

lose their credibility in any other collaboration" if SERC reneges on the EURO-GAM agreement.

John Sharpey-Schafer, from the University of Liverpool, said that the CERN subscription must be separated from the rest of the Nuclear Physics Board's spending if British nuclear-structure research is not to be "killed" by the proposed cuts. Sharpey-Schafer aims to get questions asked in the House of Commons to highlight the physicists' case.

The one ray of hope last week was that, by devising a package of budget cuts threatening an entire field of research, SERC may be better placed to win extra support in next year's science budget. "If Richmond is playing a cunning game to get more money", said Sharpey-Schafer, "I want [him] to win".

But others fear the government would call SERC's bluff in such a case. John Lilley, who heads the nuclear structure group at Daresbury, stressed that physicists' jobs are on the line.

British nuclear structure researchers are being held at gunpoint and waiting for the trigger to be pulled, he said before the meeting.

Peter Aldhous

Gulf war clouds convention

Washington & London

As international negotiating teams arrive in Washington this week to begin hammering out what could be the most ambitious international environmental treaty ever, hopes for smooth sailing have been dimmed by news that many developing nations may not attend. Also absent may be President George Bush, the host, who may be forced by the Gulf conflict not to deliver his planned opening address, that has been eagerly awaited for signs of possible changes in US policies on global warming.

Even without added international tensions, next week's meeting of the Intergovernmental Negotiating Committee on a Framework Convention on Climate Change was going to be a difficult balance of diplomacy and science. Organized by the United Nations, it is meant to build from last year's assessment Intergovernmental Panel on Climate Change (IPCC) and produce a framework convention.

The meeting starts negotiations towards a treaty first to cut emissions of greenhouse gases worldwide and then to reverse their increase in recent decades. Participants will have to agree to tolerate considerable scientific and economic uncertainty — which the United States has so far resisted — and to accept painful changes and even cuts in energy use.

The possible absence of many developing nations next week will complicate the process. A UN General Assembly resolution at the end of last year set up a small trust fund to assist delegates from poor countries and the small island states most threatened by a rising sea level. But in the short time since and because of the distractions of war, few contributions have yet been made. The organizers fear that it is now too late for some countries to attend, and that the disruption of commercial airline flights around the Middle East can only make things worse.

Procedural issues are expected to dominate the ten-day negotiating session from 4 to 14 February. Although UN officials are trying to anticipate the administrative difficulties, they acknowledge that fractious debate over the membership of working groups and even the seating arrangements could consume the entire meeting, especially if the developing countries believe themselves to be under-represented.

Strong leadership would help, but may be in short supply. Although the United States is hosting the convention, US officials stress that they are participants like any others and will not try to run the show, a position taken to avoid charges of bullying such as have occasionally arisen in previous international environmental negotiations.

The organizers hope to fill the leadership vacuum by appointing an international committee to steer the negotiations, but even that plan is controversial. The industrialized countries of the West favour a small steering group, with five or so members, where their interests would predominate. Developing nations have other thoughts.

Given the likely diplomatic squalls, hopes that the Washington meeting will produce an early draft of a framework climate convention seem unfounded. Although there is supposed to be a document for signature at the 1992 UN Conference on Environment and Development, to be held in Brazil, no country is likely to table a formal proposal until a negotiating structure is established.

Next week's meeting may accomplish that, at least. Working groups will be set up to negotiate individual sections of a convention, to deal with the legal problems and with the transfer of technology to developing countries. Some participants plan to argue for a parallel set of working groups to agree on prompt cuts in carbon dioxide and other greenhouse gases, but are likely to be opposed by the powerful US delegation.

Given the success of the Vienna Convention on the protection of the ozone layer and the Montreal Protocols attached to it, a similar structure is now the favoured model for a climate treaty, despite the differences between the two problems in scope and depth of current scientific understanding.

Critics of the Vienna-Montreal model argue that, compared with the probable costs of large carbon dioxide cuts, the economic costs of the CFC reductions required at Montreal are small. And, in marked contrast with current disagreements on global warming, governments and (eventually) chemical industries were largely in favour of CFC limits when the Montreal Protocols were completed.

Even so, as Pamela Wexler of the University of Maryland Center for Global Change remarks, people — and even environmental diplomats — tend to stick with what they know best. And "A lot of people made their reputations on the Montreal Protocol", she says. A new book from the World Resources Institute* collects a number of alternative models from some of the veterans of international environmental diplomacy. Whether next week's meeting will escape from house-keeping long enough to consider them remains to be seen.

Christopher Anderson & Peter Aldhous

*Greenhouse Warming: Negotiating a Global Regime (World Resources Institute, 1991).

Doors are opened to advanced study

Munich

Two new 'institutes for advanced study' were announced independently earlier this month in Budapest and Prague. Both institutes bear the name of the renowned Princeton institute, and they share the aim of slowing the alarming brain-drain in the region, but the similarities end there.

The Budapest institute, known as the 'Collegium Budapest', is meant to perform a bridging function with the rest of the scientific world and at the same time address some of the most acute problems in the region. It will focus at first on pressing problems such as the transformation of planned economies into market economies and interdisciplinary environmental research. Later, the Collegium might shift its focus more to natural sciences.

Beginning in the autumn of 1991, up to 15 scholarships a year will be made available to researchers from both Eastern Europe and the West; both established researchers and promising younger researchers will be eligible to apply.

The Collegium, which will be financed initially by a variety of Western European governments and foundations through the Ernst Reuter Foundation in Berlin, will be built up along the lines of the Wissenschaftskolleg, a 10-year-old 'institute for advanced study', which is in turn modelled on the eponymous institute in Princeton. A broad spectrum of political parties as well as the Hungarian Academy of Sciences have promised to support the Collegium, which will be centrally located in a former Carmelite cloister in the Buda area, where all researchers can work and some may be housed.

The Prague Institute for Advanced Study (PIAS), which hopes to open its doors within 18 months according to coordinator Petr Pechan, is intended to be a centre of excellence that can help Czechoslovakia hang on to its best researchers in natural and environmental sciences. PIAS will offer posts of two to three years duration to researchers in a number of fields as well as research grant money. Areas considered include biomaterials, peptide chemistry, environmental medicine and molecular biology, but the final decisions will depend on the success of recruitment.

Universities and research institutes in both the Czech and Slovak parts of the country have thrown their support behind PIAS, says Pechan, a Czech-born biomedical researcher who returned last year after 22 years abroad. They are less afraid of losing their most talented people to PIAS, he says, than they are of people leaving the country entirely. But Pechan is still trying to attract firm offers from Western organizations that could guarantee the future of the institute.

Steven Dickman

Spectre of contamination

London

NEW evidence has emerged that the CBL1 virus isolate used to develop one of the first commercial diagnostic blood tests for HIV infection may in fact have arisen from laboratory contamination of a cell culture with LAV-1 or HTLV-IIIB — the HIV isolates discovered respectively by Luc Montagnier at the Pasteur Institute in Paris in 1983 and by US National Institutes of Health researcher Robert Gallo in 1984.

The commercial implications of the discovery are not yet clear — a European patent has not yet been issued for LAV-1. But the possibility of laboratory cross-contamination, in one of the most respected HIV research laboratories, underlines a serious general problem with cell culture contamination that has been a hazard since the early days of AIDS research in the mid-1980s.

Amino acid sequence analysis shows that CBL1 is remarkably similar to both LAV-1 and HTLV-IIIB. Given that both viruses were being cultured in his laboratory, Robin Weiss, from the Institute of Cancer Research in London, cannot rule out the possibility that a cell-culture used to grow HIV isolated from a London patient in 1984 became contaminated with LAV-1 or HTLV-IIIB (see Scientific Correspondence, page 374).

Gerald Myers, who runs an international database of HIV sequences at the Los Alamos National Laboratory, says that nucleotide sequences diverging by as little as three per cent at variable regions of the *env* gene almost certainly indicate that the isolate came from the same patient. Slightly more divergence, typically 3 to 7 per cent, is found between 'epidemiologically-linked' cases, such as HIV-positive mothers and their infected babies. Genuinely unlinked isolates usually diverge by much more than 10 per cent at the *env* gene.

The new sequence data for CBL1 are for amino acid sequences coded by *env* and two other variable genes, so do not correspond precisely to Myers' data. Overall, these sequences diverge by only about 2 per cent from those of LAV-1 and HTLV-IIIB. Peter Highfield, head of molecular biology at Wellcome Diagnostics, which markets the UK HIV test, says the divergence between the isolates at the nucleotide level will be similarly low.

CBL1 is the latest in a line of virus isolates that have been found to be remarkably similar to LAV-1, suggesting that they arose from laboratory cross-contamination. The similarity between LAV-1 and Gallo's HTLV-IIIB is the most celebrated, sparking the accusation that Gallo had 'stolen' Montagnier's virus — dismissed in a subsequent NIH investi-

gation (see *Nature* 347, 603; 1990). Genetic sequences from two isolates, published in the *Journal of Virology* last year, which were thought to be from separate patients in New York and Nebraska, show they are also very similar to LAV-1 (see *Nature* 347, 3 & 18; 1990).

Contamination of cell cultures with macaque simian immunodeficiency virus (SIV) in the laboratory of Myron Essex at Harvard is also now accepted as the probable origin of isolates originally described as STLV-III_{AGM} from an African green monkey and HTLV-IV, from a woman in West Africa (see *Nature* 331, 562; 1988).

The precise route of cell culture contamination in these cases is not known — contamination via glassware is one pos-

sibility, or by virus particles in aerosol droplets entering the atmosphere of hoods used to manipulate cultures. Because viruses tend to adapt to cell culture after repeated generations in the laboratory, even small amounts of a contaminating virus can out-compete a new isolate and come to dominate a culture.

Simon Wain-Hobson, from the Pasteur Institute, says that the contamination problem should not be seen as professional failure on the part of AIDS researchers, rather as an "occupational hazard". Other areas of virology have been similarly afflicted, he says. The research climate in AIDS laboratories in the mid-1980s should also be borne in mind, according to Wain-Hobson — producing new HIV isolates was a high priority, to "get a handle" on AIDS.

Peter Aldhous

UK SCIENCE BUDGET

Allocation pleases no-one

London

ALL five British research councils say that they must postpone planned research projects, because of the disappointingly low UK science budget, which was divided between the research councils last week.

The Medical Research Council (MRC) seems to have fared best in this year's shareout (see table), announced by Secretary of State for Education and Science Kenneth Clarke. But Nick Winterton, head of the MRC secretariat, says that £4.1 million of that money is earmarked to begin restructuring British clinical research. The plan is to close the MRC Clinical Research Centre at Northwick Park Hospital in North London in 1994, establish a new centre at the Royal Postgraduate Medical School in London, and improve clinical research at a number of other UK teaching hospitals.

RESEARCH COUNCIL BUDGETS (£ million)

	1991-92	1990-91
Agricultural and Food	93	85
Economic and Social	35	36
Medical	200	185
Natural Environment	122	135
Science and Engineering	451	438

£15 million goes to the Royal Society

Leaving this money aside, the MRC is left with an increase of only 5.8 per cent over last year's allocation, lower than the current rate of inflation. New research into neurodegenerative diseases, diabetes and molecular genetics will be postponed, Winterton says, and spending in the MRC's existing research units and on research grants will fall, unless inflation drops rapidly.

The Science and Engineering Research Council (SERC)'s budget crisis (see *Nature* 347, 377; 29 November 1990) will

not be eased by its allocation for 1991-92, described as "lousy" by SERC chairman Sir Mark Richmond. On 6 February, SERC's governing council will produce a detailed package of spending cuts, postponing several large new projects and cutting back existing research.

The other three research councils are not planning cutbacks in their existing programmes this year, but are shelving some new research. The Agricultural and Food Research Council says that a joint clean technology initiative with SERC (see *Nature* 346, 691; 23 August 1990) will be restricted to a "modest start", and research on the biological impact of global environmental change must be delayed.

The Natural Environment Research Council (NERC)'s allocation is less than last year's, but much of the 1990-91 spending was earmarked for the Antarctic research ship *James Clark Ross*. NERC secretary Eileen Buttle says the NERC is now in the "eye of a storm", and will be forced to cut back if more money is not available next year.

The Economic and Social Research Council says that new research into global environmental change and the evolving economic and social climate of Eastern Europe will go ahead, but the number of new projects will be smaller than hoped.

Traditionally, the detailed advice on the allocation of the science budget given by the Advisory Board for the Research Councils (ABRC) to the Secretary of State has been released with the research councils' budgets. But that openness — a rare blemish in the strict policy of secrecy over the British public expenditure process — has been discontinued. In a terse statement, Clarke simply said he had followed the ABRC's advice.

Peter Aldhous

Japan's project stalls

Tokyo

AMBITIOUS plans by Japan's scientists to launch a human genome project have foundered in the cash-starved Ministry of Education, Science and Culture (MESC). MESC will set aside a small amount of money this fiscal year to establish the first division of a new human genome analysis centre at Tokyo University's Institute of Medical Science and grants for genome research will be increased. But the sums of money involved are considerably less than the modest amount scientists were hoping for. Failure to launch a large-scale project is likely to lead to renewed US criticism of Japan's failure to contribute sufficiently to the international effort to map and sequence the human genome.

Last summer, five MESC subcommittees of nearly 70 scientists led by Kenchi Matsubara of Osaka University compiled a report proposing that over ¥6,000 million (\$45 million) be spent on a five-year project. But, even before the report was printed in October, MESC officials trimmed the proposal down in budget requests for 1991 submitted at the end of August. Now those requests have been further slashed in end-of-year negotiations with the Ministry of Finance.

The report recommended spending ¥5,000 million for research grants and ¥650 million to create 35 new posts for genome research at about five research institutes and universities around the country over a five-year period. It also proposed setting up the new human genome analysis centre at Tokyo University over the same five-year period with 36 new researchers in eight divisions.

But in MESC's budget request, only ¥500 million was allocated for grants in 1991, half the average annual amount suggested. This has been further reduced to ¥400 million during finalization of the budget, according to some of the scientists involved in compiling the October report.

Only two new posts will be created around the country each year rather than the seven suggested. And the number of divisions at the new centre at Tokyo University has been cut from eight to three.

The first, the 'genome data base division', will open this year with four researchers — a professor, an associate professor, a research assistant and a technician. But, although MESC scientists were hoping to get about ¥100 million for facilities for the new division, a professor at the institute says that they have been told they cannot expect more than "10–20 per cent" of that figure. Furthermore, plans to rent a supercomputer for about ¥300 million a year have been shelved until next year. And the institute has been forced to look for new non-MESC sources of funds to support the new division.

Money is not the only problem for the new centre. The Institute of Medical Science has no biologists with expertise on computing and heated discussions are now under way to decide whom to appoint as professor of the new division, a process which is done behind closed doors rather than through open advertising for applicants.

MESC scientists originally hoped that the new centre would be established at the National Genetics Institute in Mishima where the small DNA Data Bank of Japan (DDBJ) is located. But some researchers at Mishima opposed the idea because they did not want to get involved in providing DNA analysis services for all of Japan.

MESC is not the only government organization supporting genome research. There are small genome-related research projects supported by the Science and Technology Agency, the Ministry of Health and Welfare and the Ministry of Agriculture, Forestry and Fisheries. But rather than helping the launch of a project, the large number of government organizations involved has probably been a hindrance.

Discussions are now under way between the various ministries and agencies on how to "cut the pie" of genome research, says Kennichi Matsubara. But Japan's only hope for a substantial human genome project now seems to lie with the Ministry of International Trade and Industry (MITI) and private companies.

David Swinbanks

ENVIRONMENTAL RELEASE

Tomatoes approved

Tokyo

AFTER a tortuous bureaucratic process, Japan's Ministry of Agriculture, Forestry and Fisheries (MAFF) last week approved Japan's first release of a genetically engineered organism into the open environment.

This month the ministry's National Institute of Agro-Environmental Sciences in Tsukuba will begin field experiments with tomato plants that have been genetically engineered to resist tobacco mosaic virus. The tomato plants will be grown alongside unaltered plants in open fields surrounded by only a fence and trees to investigate natural pollination and growth of the plants, and their effects on soil bacteria and flora.

Approval was given only after closed experiments had been carried out in the laboratory and greenhouses for over two years under two sets of regulations for the handling of genetically engineered organisms laid down by MAFF and the Science and Technology Agency. David Swinbanks

Polytechnics seeking their fair share

London

RESEARCHERS at Britain's 32 polytechnics are disappointed that this year's budget for the research councils (see page 359) has not been increased by the planned transfer of £50–£100 million from the budget of the Universities Funding Council (UFC). That transfer of funds, intended to cover overhead costs, has now been delayed until 1992–93 (see *Nature* 348, 183; 15 November 1990).

The transfer will help redress the imbalance of research support for the two sectors of British higher education. But many in the polytechnics now say that wholesale reform of British research support is needed if they are to compete on even terms for research council grants.

Under the traditional British 'dual-support' system, some £800 million of UFC spending on the universities (about one-third of UFC's annual budget) is reckoned to underpin research, financed by more than £200 million a year in direct grants from the research councils. By contrast, the polytechnics' £6.7 million research council income in 1988–89 was supported by £30 million provided separately from recurrent costs by the Polytechnics and Colleges Funding Council (PCFC). (Last year, a PCFC-appointed committee recommended that polytechnic research spending should be increased to £58 million in 1991–92 — see *Nature* 347, 217; 1990.)

The planned transfer, suggested by the Department of Education and Science last year, would also make the research councils responsible for the overhead costs of research — equipment maintenance, library costs and the like (*Nature* 343, 199; 1990). But difficulties in identifying these costs have delayed the transfer by a year; the research councils and the Committee of Vice-Chancellors and Principals (CVCP) now plan to produce a detailed cost-analysis for a sample of university research projects.

The transfer of UFC funds to the research councils was designed to improve the management of public research spending in universities rather than to help the polytechnics. One difficulty is that the estimate that one third of UFC spending supports research is notional only, that it includes a substantial salary element and that much of this may now be less applicable to research than in the past, given the general pressure on university budgets.

Terence Burlin, rector of the Polytechnic of Central London, nevertheless argues strongly that the vast majority

US RESEARCH COSTS

Showdown over overheads?

of the UFC's research support should be transferred to the research councils.

Money would then reach the best researchers, he says, irrespective of whether they work in universities or polytechnics. This would mean a reduction in some universities' incomes, but Burlin says none should be faced with worse than a budget shrinking at five per cent a year — a rate of contraction already experienced by a number of polytechnics.

This view is not confined to the polytechnics. John Ziman, director of the Science Policy Support Group, an independent body dependent on research council support, says the current distribution of research funds confuses British science policy.

He remarks that there is no audit of the UFC's research spending after it reaches the universities, so that UFC research money may be diverted into teaching, for example.

Some research council officials say privately that further transfers from the UFC's budget are possible now the process has begun. But switching money from the UFC is not the only way to direct more public research spending into the polytechnics. Another approach would be to merge UFC and PCFC, allowing UFC's research funds to be distributed across all British higher education institutions.

The British Labour party has made the relaxation of the 'binary divide' a policy goal. Higher education spokesman Andrew Smith says that PCFC and UFC would be merged "in the early part" of a Labour administration, but that the allocation of research money by the merged body has yet to be determined.

Closer links between PCFC and UFC are likely whatever the colour of the next government; the two councils move into the same building in Bristol later this year. But even if there is a full merger, the pressure for value for money may focus research spending into fewer institutions, meaning that few polytechnics would benefit, and that university budgets would shrink.

Yet further reductions in the universities' research income from the UFC will meet strong resistance. John Ashworth, director of the London School of Economics and chairman of a joint CVCP/Committee of Directors of Polytechnics working group on higher education finance, says that decisions on research management are best made in the institutions where research is carried out. Transferring the majority of the UFC's research spending to the research councils would result in funds being allocated "with the skill and panache of the Soviet Academy of Sciences", he says. He believes the polytechnics should concentrate on providing high-level training, rather than aspiring to be "second-class universities".

Peter Aldhous

Washington & San Francisco

THE glory days of university research in the United States, when the government paid universities to build hundreds of laboratories and keep them running, may be at an end. And the "indirect costs" of academic research are becoming a divisive issue between universities and the government and even within universities.

After decades of reimbursing universities fully for indirect costs, which may include everything from new construction to heating and upkeep, federal science agencies are increasingly paying for just the science and a fraction of related costs. University officials say the trend may turn academic scientists against their administrators and eventually cause the collapse of university research.

Budget pressure and congressional legislation last year forced both the National Science Foundation and the US Department of Agriculture to reduce payments of indirect costs. Agriculture was told that payments should not exceed 14 per cent of direct research costs, about a quarter of what the average university calculates as its real overhead expense. Now the White House has signalled that the National Institutes of Health (NIH) may have to cut back as well.

The NIH are the largest single source of funds for university-based research and one of the few science agencies still paying full indirect costs. In an internal memorandum to NIH, the White House Office of Management and Budget (OMB) last month sanctioned less than full reimbursement of indirect costs in NIH-funded university research.

The document is NIH's spending authority for the year, saying how much they can spend, and how. Although the language is outwardly harmless — "Indirect costs are upper limits", for example — both the academic community and NIH are on full alert.

At an advisory council meeting of the National Institute of General Medical Sciences last week, NIH extramural research director John Diggs described the OMB plan as "shocking". In a subsequent interview, he said that NIH nevertheless remain committed to full support of indirect costs, although "we're in line to have further discussion with OMB".

University representatives are preparing for the worst. Carol Scheman of the American Association of Universities says "virtually every research university has made long-term commitments in the belief that it will be reimbursed . . . the collapse of a reliable reimbursement policy would mean a major financial crisis".

University administrators warn that if federal agencies refuse to pay full indirect

costs, they may have to recover the money by requiring researchers to include the costs of support services, such as animal facilities, in direct research costs.

Researchers, in turn, worry that grant proposals bloated in this way will disadvantage them in what is already a cut-throat fight for federal research support. Speaking of an "an erosion of trust between faculty and administration", Douglas Kelly of the American Association of Medical Colleges says he is worried about "a coming confrontation".

Meanwhile, Congress has taken up the issue as an example of the misuse of government funds. Disturbed by the increase in the cost of research (the average value of NIH grants has doubled in the past decade, largely because of indirect costs), Representative John Dingell (Democrat, Michigan) has unleashed his staff investigators on Stanford University, which has one of the highest indirect cost rates at 72 per cent. The investigation has found that, among other questionable practices, Stanford improperly included in its administrative overhead payments on a university yacht and furnishings for president Donald Kennedy's residence.

Last week Kennedy announced a three-pronged initiative to clear the air. Stanford will hire an outside accounting firm to evaluate the university's accounting methods, appoint an advisory panel to examine research costs and voluntarily withdraw some \$500,000 of past claims.

Kennedy says that he does not feel that Stanford has been singled out unfairly, although it is certainly "not pleasant" to be put through such an investigation. Perhaps because of the university's high profile and some \$500 million in indirect cost charges in the past decade, "we just happened to be first", he says.

Dingell's Energy and Commerce oversight subcommittee plans to hold hearings on the subject in February or March. Other universities, including the Massachusetts Institute of Technology, may also be investigated, according to staff member Dennis Fitzgibbons. Dingell's next step, Fitzgibbons says, is to determine what changes in the system are needed. If there are restrictions on the payment of indirect costs, Congress must determine who should decide what is reasonable. NIH could set a maximum reimbursement level, the federal government could draft new regulations on indirect-cost recovery or Congress could limit indirect-cost rates in the appropriations process, as it did last year with the agriculture department.

Kennedy says he cannot predict what will be decided, but he predicts that modifications to the indirect-cost system are on the way.

**Christopher Anderson
& Elizabeth Schaefer**

Turning off the gas . . .

SIR—In the leading article "Next steps on global warming" (*Nature* 348, 181; 1990) you emphasize the need now — after the Second World Climate Conference (SWCC) in Geneva — for a more persuasive framework of arguments than the Intergovernmental Panel on Climate Change (IPCC) has provided. Clearly, governments must reach agreement on a comprehensive framework for a World Climate Convention in 1992 that equitably balances the protection of the global environment against the needs of developing countries and the world's capacity to generate wealth and thus well-being.

A beginning has been made by the final statement of the scientific/technical sessions of the SWCC. Action is needed to manage the risk of climate change and to identify rational strategies to mitigate or prevent adverse effects of expected climate changes. Such strategies might involve ways of slowing climate change that will give countries more time to enhance their prospects for a sustainable development.

The SWCC scientific/technical statement, however, does not indicate how the results of the IPCC should be amended by a critical and unbiased study of options for managing the risk of greenhouse warming. It states only that continued dialogue is necessary between scientists and policy-makers.

To help this process of development of greenhouse warming policy, the "First Nordic Inter-Disciplinary Research Conference on the Greenhouse Effect: Meteorology, Climatology, Effects, Risk Management, and Interim Climate Modification" is being organized in Copenhagen on 18–20 February 1991 with the chairman of the IPCC, Professor Bert Bolin, as programme chairman (see announcement in *Nature* 25 October 1990, Classified 28).

The conference will focus on novel approaches to managing the risks involved in the decision-makers' dilemma associated with the anthropogenic greenhouse effect. Among the options to be critically discussed will be engineered, interim climate modification, considered as a 'technological insurance', in order to gain the time required for implementing at an economically feasible pace and without disastrous environmental effects the Brundtland Commission's programme of drastically reducing the rates of release of greenhouse gases to the atmosphere.

SIGURD O. NIELSEN*

Symplexor SciTech Systems Research & Policy Analysis,
Holte, Denmark 2840

*Present address: National Center for Atmospheric Research, Interdisciplinary Climate Systems, Boulder, Colorado 80307-3000, USA

Uphill struggle

SIR—The analysis of bicycling by Franke *et al.* (*Eur. J. Phys.* 11, 116; 1990), as outlined by John Maddox (*Nature* 346, 407; 1990), assumes that the cyclist does not regulate the steering angle and lean angle in combination during motion. Unfortunately, as most cyclists are aware, these two variables are under constant adjustment in order to maintain balance, even during hands-off riding. If the steering angle is fixed, it is almost impossible to maintain one's balance. This could be taken as confirmation of the finding that there are only one or two combinations of steering angle, rider displacement and lean angle consonant with maintenance of balance and direction; the poor cyclist's control is unlikely to be so sensitive. Furthermore, it should be remembered that for the cyclist to maintain constant velocity on the flat, he must be pedalling, and this action results in a constant oscillation in the centre of gravity and displacement.

Maddox also suggested that riders stand on the pedals when going slowly in order to achieve a lengthening of the trailing distance. The only reason why I stand on the pedals is to increase the power output when climbing hills or accelerating rapidly.

JONATHAN SCOTT

Department of Physiology,
University of Leicester,
PO Box 138,
Leicester LE1 7HN, UK

One route to religious ecstasy

SIR—The hallucinogenic mushroom fly-agaric (*Amanita muscaria*) may be the oldest pharmacological agent used in religious rituals by primitive societies to attain states of trance. By means of such ecstatic experiences, individuals such as shamans and medicine-men cured the diseased, accompanied the dead to the underworld and acted as mediators between their community and its gods¹. The role of fly-agaric in the ecstatic rites of Siberian shamans is also well documented² and there is strong indirect evidence for the identification as *A. muscaria* of Soma, an inebriating plant used by the earliest Indo-European populations of India and celebrated in the hymns of the *Rigveda*³.

In mediaeval Europe, hallucinogenic plants were used, especially among country-folk, in popular religious rites that were condemned by the Church as witchcraft and heresy (for a recent discussion, see ref. 4). The names of some of these psychoactive plants have survived in the works of Renaissance natural philosophers^{5,6}, but there is no mention of the fly-agaric in them or other sources^{3,4}. A

possible exception to this silence may be found, however, in the biography of a Genoese noblewoman, Caterina Fieschi-Adorno, Saint Catherine of Genoa (1447–1510) (reprinted in ref. 7). The following passage seems to have escaped thus far the attention of both ethnopharmacologists and historians:

God, Who had taken control even of her body, wanted to regulate it, and take away from her all the wordly and human instincts. Because He wanted her to lose the taste for the food she ate, He made her keep hepatic aloe and ground agaric, so that, when she realized that some food was giving her pleasure, or suspected so, she secretly put some into it. After God had prepared this soul in such a way, He tempted her with spiritual temptations. He infused such suavity and divine sweetness in her heart that both soul and body were so full as to make her unable to stand up.

In the second half of her life, Saint Catherine of Genoa experienced a number of trances during which, she affirmed, she was able to achieve a state of mystical communion with the "Divino Amore" (Divine Love). Describing one such trance, her biographers write: "... one day she felt as if she were up in the air; her spiritual self (*parte spirituale*) desired to reach for the Heavens, and bring with her the soul; her human self (*parte umana*) desired to reach for the Earth. Finally, after a long struggle, her spiritual self succeeded. She saw visions of angels, and laughed with them".

Complex experiences such as religious ecstasy, which obviously involve multiple material and cultural factors, cannot be explained uniquely in pharmacological terms. The passage reported here does raise, however, the intriguing possibility that at least one of the means, either conscious or accidental, by which Caterina Fieschi-Adorno attained her states of religious ecstasy was by the use of fly-agaric (which can be found in the Italian Alps). In her Christian devotion, she could not suspect herself of following the steps of the Aryas ministers of Indra, the vedic God of the Sky.

DANIELE PIOMELLI

Unité de Neurobiologie et Pharmacologie
(U.109) de l'INSERM,
Centre Paul Broca,
2ter rue d'Alésia,
75014 Paris,
France

1. Eliade, M. *Le Chamanisme et les Techniques Archaïques de l'Extase* 2nd edn (Payot, Paris, 1968).
2. Furst, P.T. *Hallucinogens and Culture*, 89–108 (Chandler and Sharp, Novato, California, 1976).
3. Wasson, R.G. *Soma, Divine Mushroom of Immortality* (Harcourt, Brace and World, New York, 1969).
4. Ginzburg, C. *Storia Notturna. Una Decifrazione del Sabba* 283–288 (Einaudi, Torino, 1990).
5. Cardano, G. *De Subtilitate* Vol. 18, 354 (J. Petreium, Nuremberg) (1550).
6. Della Porta, G.B. (1558) *Magiae Naturalis, sive de Miraculis Rerum Naturalium* First ed Vol. 2, 186 (Naples, 1558).
7. Craveri, M. *Sante e Streghe* 144–161 (Feltrinelli, Milano, 1980).

Olbers' paradox has more to teach

The question of why the sky is dark at night is a textbook problem, but textbook answers may still miss the point. The problem is that of determining which is the more dominant of the two preferred solutions.

OVER the years, Olbers' paradox has become a kind of classic — it is a real paradox, which everybody knows can be resolved in an interesting way. Most often, the paradox and its resolution are advanced as evidence that the theory of relativity is not merely compatible with the night sky that we know, but that it is essential in accounting for its appearance. But even paradoxes, it appears, are not always what they seem.

The issue is simply that of why the night sky is not uniformly as bright as the surface of the Sun. That, after all, is what one would expect if the stars in the Universe number something like infinity, or at least if there are so many of them that their collective surface area projected on the sphere of the sky would suffice to cover it completely — or to tile it, in the technical sense.

Then the output of visible radiation from any element of the night sky would be more or less identical with that from an equal area of the surface of the Sun. Of course, there would be occasions when cool stars would be interposed between ourselves and brighter but more distant stars, but that would merely give the night sky a granular appearance. In general, we would have the uncomfortable impression that we live at the centre of a hollow black body whose temperature is about 6,000 degrees centigrade.

The question why the real world is not like that is Olbers' paradox. And a little reflection will show that some obvious resolutions will not suffice. If, for example, the Universe is supposed to be uniformly populated by stars like the Sun, filling the Universe uniformly with obscuring dust will not get round the difficulty. The intensity of the radiation from a distant star will decrease exponentially with its distance, but cannot do so quickly enough to resolve the paradox without also obscuring the Sun itself. And what, in any case, would be the fate of such a dust-filling in a black-body oven of this kind?

Arguments like that explain why the resolution of the paradox must be in some way cosmological. In simple language, there are two ways of doing the trick. First, if the galaxies that house the distant stars are finite in age, there has only been a finite time during which they can have been producing radiation; the radiation field in the black-body oven may be destined to reach the typical surface temperature of a star, but there has not yet

been time enough for the oven to reach equilibrium.

Second, and more appealing, if the Universe is expanding, the radiation from distant stars is shifted towards the red end of the spectrum, while the volume of the Universe that must be filled with radiation is continually and continuously increasing. These, of course, are the arguments that make Olbers' paradox a part of the case for believing in relativity.

The two arguments are, of course, distinct, even though they may be linked together in some views of what the Universe is like. So which of them should carry the greater weight? Paul S. Wesson from the University of California at Berkeley and from the University of Waterloo at Toronto has for some years been conducting a personal crusade to rid the world of erroneous resolutions of the paradox. His latest essay in this direction, in the 1 February issue of the *Astrophysical Journal* (367, 399; 1991) reiterates the conclusion of his earlier arguments that it is the finite age of the galaxies, and not the expansion of the Universe, that chiefly accounts for the resolution of the paradox.

The question of which effect is the more important is necessarily quantitative. Both effects must have some influence; the issue is that of which influence is the greater. The result is that the finite age of the galaxies accounts for a reduction by a factor of ten in the intensity of intergalactic radiation (which Wesson calls extragalactic background light), and that the relativistic effects, notably the expansion of the Universe, account for a factor of only three or thereabouts.

Amusingly, Wesson starts with an account of how the textbooks handle (or mishandle) Olbers. Most of the textbooks, he says, are "still less than satisfactory" even though the "situation is better now". Wesson's article, in the circumstances, may properly be regarded as a kind of book review on one technical issue, but the paper also has a calculation of the extragalactic radiation field to be expected with different models of the expanding Universe. He writes with something of the resignation of one saying the last word on a topic for the umpteenth time.

But how does one set about calculating the expected radiation field? It is not, of course, a trivial problem. Allowing for the shifting of the radiation towards the red

end of the spectrum is easily accomplished for a particular model of the Universe, but it is also necessary to allow for the mode of evolution of the galaxies, which are generally supposed to have evolved from hotter to cooler, but in some way that cannot easily be determined. (If it were otherwise, the big problems of cosmology would surely have been solved a long time ago.)

For Wesson's purpose, that of calculating an order of magnitude, it is sufficient to suppose that the average temperature of a galaxy is a decreasing function of time, and to use that as a basis for calculating the output of radiation from a galaxy using the familiar Planck formula for the spectral intensity of the radiation from a black body.

With the help of some remarkably straightforward algebra, Wesson derives an expression for the light intensity from an ensemble of galaxies which depends on the two effects with which he is principally concerned in distinct ways. The finite age of the galaxies enters by means of the lower limit of an integral over time, while the strictly relativistic effects arise through a scaling factor determined by the relativistic model of the expansion.

The virtue of this piece of algebra is that it makes it possible to compare the sensitivity of the radiation to the two effects — the finite age of the galaxies and the relativistic expansion — without having to be numerically sure about all of the other details.

In the end, the comparison is made even less dependent on the values of irrelevant parameters by the device of directly comparing the intensity of the intergalactic radiation field in an expanding and a static model of the Universe. What Wesson finds is that the ratio rarely exceeds 1:3, and is often much less. By contrast, there are numerical arguments to suggest that the effect of the finite galactic age will be greater.

In the circumstances, it hardly seems appropriate that people should go on telling students that the resolution of Olbers' paradox is a relativistic effect. It is more relevant that there was a time in the history of the Universe when there were no galaxies of the form in which we know them, and thus no opportunity to fill the Universe with radiation of the kind now called visible. What happened before then, of course, remains anybody's guess.

John Maddox

A stop-start ocean conveyor

Glenn A. Jones

It is becoming clear that the end of the last glacial period was not a simple affair. Lasting from 15,000 to 8,000 years before present (BP), the last deglaciation was characterized by a series of up to four climatic oscillations, during a time when changes in solar heating (due to orbital variations) were limited and gradual. Several hypotheses have been put forward to explain this complicated system, only to be challenged, modified or discarded. The latest hypothesis, from Broecker and colleagues¹, is that natural oscillations in the salinity of the North Atlantic alter the ocean's large-scale circulation, and hence its influence on the climate.

Before this latest contribution, Broecker and colleagues²⁻⁴ had advanced an elegant explanation of why the climate of the North Atlantic returned briefly to nearly full glacial conditions about 11,000 years ago during an interval known as the Younger Dryas. The hypothesis required that meltwater from the thawing Laurentide Ice Sheet on North America be diverted from the Mississippi River (whence it flowed into the Gulf of Mexico) to the St Lawrence River, to flow into the North Atlantic. Evidence was found in carbon-14 dated cores from the Gulf of Mexico that meltwater discharge from the Mississippi River did cease during the Younger Dryas, and glacial geological evidence suggested a concomitant dis-

charge down the St Lawrence.

Unfortunately, no well-dated ocean cores are yet available that unambiguously demonstrate that significant amounts of meltwater were discharged directly to the North Atlantic via the St Lawrence. Nevertheless the argument was made that meltwater discharge into the northern North Atlantic was sufficiently large to prevent deep water formation (see box), a process which today releases some 5×10^{21} calories of heat to the atmosphere over the North Atlantic each year. This reduced heat flux halted the retreat of the Northern Hemisphere icesheets and led to a return to near glacial conditions around the North Atlantic.

In 1989, Fairbanks demonstrated⁵ that the rate of sea level rise during the last deglaciation, an indicator of the rapidity of net icesheet decay and meltwater release to the world oceans, was enhanced just before and immediately after the Younger Dryas interval. Thus, the Younger Dryas interval itself was shown to be a time of reduced net meltwater discharge. This evidence appeared to discredit the diversion hypothesis, as the reduction in meltwater discharge through the Mississippi River could be explained by a reduction in meltwater production overall rather than a diversion, and led to Shackleton's sapient prediction in *News and Views*⁶ that Broecker would soon

erect an even more elegant hypothesis in his next contribution.

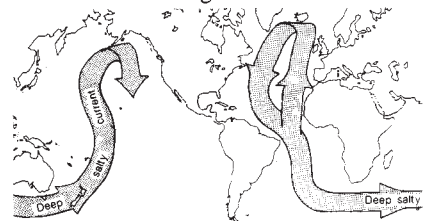
As predicted, Broecker and colleagues have acknowledged¹ the weakness of the diversion hypothesis and offered a much improved explanation of the expanding body of high-quality data that has been assembled in the past few years. Most importantly, whereas the diversion hypothesis offered an explanation for only the Younger Dryas, their new natural salt-oscillator hypothesis (see box) offers a plausible explanation for the multitude of glacial-aged climate oscillations with approximate 2,000-year periods that are seen throughout the interval 20,000–60,000 BP in Greenland ice cores⁷, and in high-deposition-rate cores from the North Atlantic^{7,8}. The last deglaciation was itself characterized by up to four intervals of accelerated meltwater discharge into the North Atlantic with an approximate 2,000 year cyclicity⁹. Each of these events resulted in a shutdown of the 'thermohaline' conveyor¹⁰. Even during the more recent past with much-reduced ice cover, there is evidence¹¹ for an alternating expansion and contraction of alpine glaciers in Sweden, Alaska and the Yukon with a recurrence of about 2,500 years.

If these oscillations, which all appear to have a frequency of approximately 2,000 years, are real and they were forced within the climate system, not by external variations, it is logical to implicate the ocean as it operates with a time constant close to that of the observed oscillations. Following up the conceptual model¹ of a salt oscillator, Birchfield and Broecker¹² performed a scale analysis of the various salinity fluxes of the system and by integration of simple salt conservation equations found that model oscillations with a period of a few thousand years occur over a wide range of salinity fluxes. Although not quite a 'grand unified theory' of climate change, the hypothesis does link, at least conceptually, the atmosphere, ice sheets, surface ocean and deep ocean via salt, heat and vapour transport with realistic fluxes and time constants.

Despite its appeal, there are at least three issues that need addressing before the model can be accepted. First, detailed, well-dated records of surface and deep-ocean conditions and of atmospheric temperature must be obtained. The records need to be sufficiently long to show that the salt-oscillator model is statistically robust, and that the phases between changes in these three components of the model are correct. At present, age control is insufficient for most data, and the highest-quality data used to support the salt oscillator come from the temporally restricted deglacial interval where there appear to be four oscillations during a time when the climate response to a salt oscillator should be amplified¹³ and more easily detected. However, with only four

Thermohaline circulation

Broecker and colleagues' ideas about the climate variations during the last deglaciation revolve around the thermohaline circulation (or 'conveyor belt') of the oceans (see figure). Currently, warm surface water moving northwards in the



The thermohaline circulation through the Atlantic and Pacific Oceans. (From ref. 15.)

North Atlantic evaporates, becoming increasingly saline, and cools. The resulting increase in density causes it to sink, forming cold bottom water. This process annually releases some 5×10^{21} calories of heat into the atmosphere over the North Atlantic. According to Broecker and colleagues' previous idea, the sudden arrival of diverted meltwater from the Laurentide Ice Sheet in the northern North Atlantic reduced the salinity and halted the

thermohaline circulation and hence the heat flux to the high-latitude atmosphere.

Natural salt-oscillator

Broecker and colleagues now argue that switching between the on and off states of the thermohaline circulation is driven by natural oscillations in the salinity budget of the Atlantic ocean. In the 'off' mode, Atlantic salinity increases relative to the remaining ocean owing to a surface salinity flux caused by a net loss of water vapour from the Atlantic to the Pacific drainage basins. When a salinity threshold maximum is reached the conveyor is turned on, driven by the increased density difference between the Atlantic and the remainder of the world ocean. With the conveyor on, there is increased surface transport of heat and salt to the high-latitude Atlantic but also increased export of salt from the Atlantic to the remainder of the ocean. The latter drains the Atlantic of salt until the salinity reaches a threshold minimum, at which the density difference between the Atlantic and the remainder of the ocean is insufficient to maintain the outflow and the conveyor is turned off again.

quasicyclic events it is not clear whether these are a manifestation of the salt oscillator or merely some random climate response. Examination of isotope stage three (approximately 20,000 – 60,000 BP) will offer the best test of the hypothesis, as the interval can be dated directly by carbon-14 dating, is long enough to contain a sufficient number of cycles for reliable statistical analysis, and meets the hypothesis requirement for ice cover on the surrounding landmasses.

The second issue is the need to define more precisely the timing and geographical distribution of glacial meltwater inputs to the North Atlantic and their effect on surface water salinity. Because deep waters are produced in particular regions in the North Atlantic, the size and location of these meltwater inputs place important constraints on whether the oscillator is enhanced or attenuated. Although coherent patterns in the timing and distribution of meltwater discharge during the last deglaciation have been identified qualitatively⁹, there is still a poor quantitative understanding of how much the salinity of the surface North Atlantic was affected by these discharges.

The third issue is the need to use a more realistic model to demonstrate that internal salt oscillations can arise naturally given plausible initial conditions. One of the admitted weaknesses of Birchfield and Broecker's model¹² is that it is driven solely by mean ocean density differences between the Atlantic and the other oceans, neglecting the important local effects which produce changes in the vertical stability of the North Atlantic. The outcome of their current model is some-

what forced in that fluxes are prescribed instead of freely predicted.

By covering a host of climate oscillations, not just the Younger Dryas period, the salt-oscillator hypothesis should usefully expand the horizons for climate researchers. With current concerns about the climate's sensitivity to anthropogenic changes, such natural oscillations and hair-trigger responses take on a profound significance. Indeed, recent observations¹⁴ show that even small-scale changes in surface-water salinity can cause measurable changes to the production of bottom water. And although the salt-oscillator model, as currently presented, requires ice-cover on the adjacent continents, it may in fact be persisting now in weakened form and could be implicated in the Little Ice Age of 400–200 years ago. □

Glenn A. Jones is at Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA.

1. Broecker, W.S., Bond, G. & Klas, M. *Paleoceanography* **5**, 469–477 (1990).
2. Broecker, W.S., Peteet, D.M. & Rind, D. *Nature* **315**, 21–26 (1985).
3. Broecker, W.S. *et al.* *Paleoceanography* **3**, 1–19 (1988).
4. Broecker, W.S. *et al.* *Nature* **341**, 318–322 (1989).
5. Fairbanks, R.G. *Nature* **342**, 637–642 (1989).
6. Shackleton, N.J. *Nature* **342**, 616–617 (1989).
7. Broecker, W.S. *Paleoceanography* **5**, 457–467 (1990).
8. Boyle, E.A. & Rosener, P. *Palaeogeog. Palaeoclim. Palaeoecol.* **89**, 113–124 (1990).
9. Jones, G.A. *Norsk Geologisk Tidsskrift* (in the press).
10. Keigwin, L.D., Jones, G.A., Lehman, S.J. & Boyle, E.A. *J. Geophys. Res.* (in the press).
11. Denton, G.H. & Karlen, W. *J. Quat. Res.* **3**, 155–205 (1973).
12. Birchfield, G.E. & Broecker, W.S. *Paleoceanography* **5**, 835–843 (1990).
13. Lehman, S.J., Jones, G.A. & Keigwin, L.D. *EOS* **71**, 1359 (1990).
14. Dickson, R.R., Meincke, J., Malmberg, S.-A. & Lee, A.J. *Prog. Oceanog.* **20**, 103–151 (1988).
15. Broecker, W.S. *Nature* **328**, 123–126 (1987).

Why the eye doesn't shape up

Brian J. Rogers

THE horizontal separation of our eyes by about 6 cm means that the positions of corresponding images on the two retinæ are typically not the same¹. These differences in position, or disparities, provide information about the relative location of objects in three-dimensional (3D) space. The horizontal separation of the eyes dictates that disparities are principally horizontal, although there can also be small differences in the vertical positions of corresponding images. It has been shown formally that these vertical disparities could provide information about the absolute distance from the observer to the viewed object, which is an attractive possibility because it would mean that the visual system would not have to rely on information from the oculomotor system^{2,3}. But the novel results reported by Cumming *et al.* on page 411 of this issue⁴ show quite conclusively that the human

visual system is not able to use vertical disparities in computing 3D shape.

When the eyes are fixated on a point in space, the horizontal disparity between the binocular images of the point is necessarily zero. As the fixation point can be either close to the observer or far away, horizontal disparities cannot provide information about the absolute distance to points in 3D space, but only about relative position. Indeed, horizontal disparities alone cannot even specify the magnitude of the relative distance (or depth) of two points on the same object or surface, because the magnitude of the disparity between the two points will vary with the absolute distance according to a simple inverse square law. Consequently, it has been assumed that horizontal disparities need to be scaled by some estimate of absolute distance in order to account for the fact that we are able to match the

amount of depth in 3D surfaces at different distances with reasonable accuracy — the phenomenon known as depth constancy^{5,6}. An obvious source of absolute distance information is that provided by the vergence angle of the eyes and there is evidence that manipulating this angle while keeping other factors constant does affect the amount of perceived depth.

In theory, vertical disparities can also provide information about the absolute distance of a surface (D) as well as the degree of eccentricity of the observer's gaze (G). The second point can be appreciated intuitively from the fact that when the eyes are gazing off to one side, the image of the fixated object will be uniformly larger in the closer eye than in the other (a in the figure). That vertical disparities can provide information about absolute distance can be understood by considering the retinal images created by a rectangular fronto-parallel surface. When the surface is close to the observer, the images in the two eyes will be slightly trapezoidal in shape because the surface is being viewed obliquely from two slightly different positions (ref. 7; b in the figure). On the other hand, when the surface is far away, the two eyes' images will both be rectangular (c in the figure). So if the human visual system could extract these small differences in vertical size, it could in theory use them to provide information about (1) the absolute distance of objects from the observer and (2) the angle of eccentric gaze. In addition, the estimate of absolute distance could be used to scale horizontal disparities and thereby achieve depth constancy (3).

What evidence is there that the human visual system is able to exploit vertical disparity information? With respect to prediction 1, no one has ever reported that appropriate manipulations of vertical disparities in stereograms have a direct effect on our perception of the absolute distance to a surface. Mayhew and Longuet-Higgins⁷ have cited Ogle's induced effect as evidence that the visual system uses vertical disparity information to estimate angle of eccentric gaze (prediction 2). In the induced effect, a vertical magnification of just one eye's image produces the impression of a surface slanting out of the frontal plane even though there are no horizontal disparities between corresponding points. If the visual system exploited the vertical disparities present, these could be used to estimate (incorrectly) an eccentric gaze angle. This estimate, in turn, could be used to reinterpret the absence of horizontal disparities (as required by the geometry of the situation) and thereby account for the perceived surface slant. Koenderink and I, on the other hand, have shown that Ogle's induced effect is also predicted if the visual system estimates surface slant directly from the amount of local defor-

Mice and men

MULTIDRUG resistance is a major barrier to cancer chemotherapy, and the evidence points to the source of the problem in many cases being overexpression in tumour cells of the so-called P-glycoprotein. This membrane protein, the product in humans of the *MDR1* gene, is a transmembrane transporter which facilitates extrusion of a number of different anti-cancer drugs from the cells in which it is expressed. G.H. Mickisch *et al.* now report (*Proc. natn. Acad. Sci. U.S.A.* **88**, 547–551; 1991) that they have made transgenic mice expressing the human *MDR1* gene in their bone marrow, and that the leukocytes of the mice are indeed resistant to anti-cancer drugs. So these animals provide what could be an ideal system for testing ways in which the P-glycoprotein can be prevented from working.

Winding up

DIFFERENT mechanisms are responsible for cyclones to the east and west of Mexico, according to J.A. Zehnder and R.L. Gall (*Tellus* **43A**, 25–36; 1991). In the Gulf of Mexico and further to the east, cyclones arise from temperature contrasts over the land and sea. But these anomalies do not occur to the west of Mexico, in the eastern North Pacific. Instead the authors argue that the Sierra Madres mountain chain, which runs northwest to southeast in the western United States, is responsible. Zehnder and Gall's model shows that the cyclonic seeds are shed by southeasterly winds, blowing parallel to the mountain range. The winds must be neither too strong nor too weak for the cyclonic vortices they generate over the Pacific to be of the right scale to draw heat efficiently from the warm ocean to develop into full-blown cyclones. This mechanism explains why cyclones tend to develop close to the Mexican coast in the Pacific; by contrast, they are distributed widely over the Gulf of Mexico and Caribbean.

Crystal tips

THE search for undue human exposure to lead turns to the drinks' cabinet with a report by J.H. Graziano and C. Blum of Columbia University in *The Lancet* (**337**, 141–142; 1991). The authors found that the lead concentration of port kept in lead crystal decanters for four months rose from the $89 \mu\text{g l}^{-1}$ measured when it was fresh from the bottle to as much as $5,331 \mu\text{g l}^{-1}$ — that is, lead had been eluted from the decanter into the port. In a further analysis, lead concentrations of the contents of decanters provided by colleagues were tested; in one instance, brandy that had been stored for over five years yielded a value of $21,530 \mu\text{g l}^{-1}$. The implication of the study is that lead-free crystal may lack sparkle but that it is less of a potential hazard to health than the leaded variety.

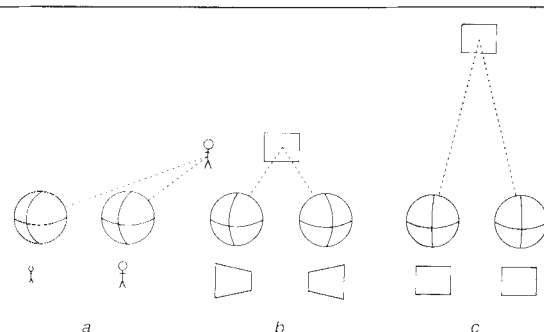
mation or shear between two stereoscopic images⁵. Thus the evidence for the second prediction is inconclusive.

With hindsight, it is surprising that it has taken nearly eight years for anyone to test the third main prediction — that vertical disparities could be used to scale horizontal disparities to achieve depth constancy. In their report Cumming *et al.*⁴ describe the results of an experiment in which

ing distance in their set-up. The authors interpret these findings as strong evidence that the human visual system does not use vertical disparities to estimate absolute viewing distance and to scale horizontal disparities. Additional support for their conclusion comes from our own unpublished results and those of E. Sobel and T. Collett (personal communication) which show that manipulations of vertical

disparity do not affect the amount of perceived depth in either 3D sinusoidal corrugations or a simple step change in disparity respectively.

Under normal viewing conditions, our ability to judge the relative depth of objects at different distances is quite good, at least for objects within a few metres of the observer^{5,6}. If disparities are not scaled by vertical disparity information, what factors are responsible for depth constancy in normal viewing? In their experiment, Cumming *et al.* show that extraretinal cues in the form of vergence angle manipulations have a "substantial" scaling effect, although this accounted for only 25 per cent of that required for complete constancy. The degree of constancy was still low (less than 50 per cent) when the authors altered the physical viewing distance (rather than just the vergence angle), suggesting that shape constancy is not well maintained by the visual system. Alternatively, the answer might lie in the way



a, When the eyes are directed towards an object which is not directly in front of the observer, the image of that object will be uniformly larger in the one eye than the other, because one eye (in this case the right eye) is closer to the object. This difference in image size increases with the eccentricity of gaze and therefore could provide a clue to gaze angle, although the size difference also depends on the object's distance from the observer. b, With fixation on a rectangular surface directly in front of the observer, the images on the retina will be trapezoidal in shape rather than rectangular. This arises because the left-hand edge of the surface is closer to the left eye than the right-hand edge and, consequently, it will subtend a larger angle. c, For the same surface, much further away, the size difference is minimal and both images will be approximately rectangular in shape. Hence, in principle, vertical size differences could be used to provide information about the absolute distance of the surface from the observer, as was originally suggested by Mayhew and Longuet-Higgins². (In reality, the retinal images would be inverted, left–right reversed and their shapes would be distorted by the curvature of the retina.)

observers were asked to judge the perceived shape of horizontal cylinders with an elliptical profile defined by disparity information alone. By using a series of cylinders which were more elongated or flattened in the third dimension, the authors were able to determine the point at which the cylinders appeared to have a circular profile, like the side view of a length of pipe. Given the inverse-square law relationship between disparities and absolute distance, the vertical disparity hypothesis would predict that a simulated cylindrical surface with the same disparity profile should be perceived as having a different amount of depth and thus a different shape when vertical disparities appropriate to different viewing distances are introduced. But Cumming *et al.* show conclusively that manipulations of vertical disparity have no effect on the perceived shape of the cylinders.

Manipulations of vergence angle, on the other hand, did alter perceived shape to the extent of 45 per cent of the constancy produced by a real change in the view-

constancy was calculated. The authors have assumed that the same estimate of absolute distance is used to scale both disparities and angular size for their shape judgements, which may not be the case. If constancy were calculated on the basis that only disparities are incorrectly scaled, then the degree of constancy would be much higher and closer to that found in other studies. □

Brian J. Rogers is in the Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, UK.

1. Wheatstone, C. *Proc. R. Soc.* **128**, 371–394 (1838).
2. Mayhew, J. E. W. & Longuet-Higgins, H. C. *Nature* **297**, 376–379 (1982).
3. Gillam, B. & Lawergren, B. *Percept. Psychophys.* **24**, 121–130 (1983).
4. Cumming, B. G., Johnston, E. B. & Parker, A. J. *Nature* **349**, 411–413 (1991).
5. Ono, H. & Comerford, J. *Stability and Constancy in Visual Perception* (ed. Epstein, W.) 91–128 (Wiley, New York, 1977).
6. Sedgwick, H. in *Handbook of Perception and Human Performance* Vol. 1 (eds Boff, K. R., Kaufmann, L. & Thomas, J. P. L.) 21.1–21.57 (Wiley, New York, 1986).
7. Frisby, J. P. *Nature* **307**, 592–593 (1984).
8. Rogers, B. J. & Koenderink, J. J. *Nature* **322**, 62–63 (1986).

Remembrance of things past

Andrew W. Murray

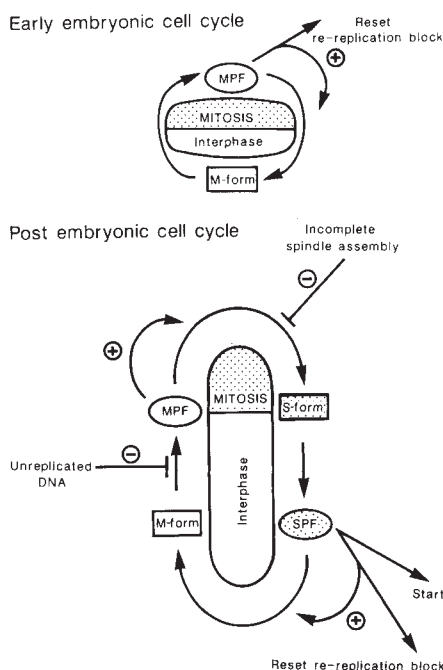
MOST eukaryotic cell cycles produce two cells that are genetically identical to their mother. Success in this endeavour requires that rounds of chromosome replication (S phase) and chromosome segregation (mitosis) alternate with each other. On page 388 of this issue¹, Broek *et al.* reveal that $p34^{cdc2}$, the ever-popular product of the *cdc2* gene, is the seat of a cellular memory that ensures this regular alternation of replication and segregation.

The protein $p34^{cdc2}$ is the central component of a biochemical engine that controls the cell cycle in all eukaryotic cells (reviewed in refs 2 and 3). In combination with the mitotic cyclins, it can form an active protein kinase, known as growth-associated H1 kinase or maturation-promoting factor (MPF), that induces the events of mitosis. Alternatively, $p34^{cdc2}$ can combine with different cyclin molecules, known as G1 cyclins, and catalyse passage through Start, an event that commits the cell to initiating DNA replication⁴. By analogy to MPF we can think of the complex of $p34^{cdc2}$ and G1 cyclins as being a Start-promoting factor (SPF). Thus the alternation of replication and segregation can be thought of as the alternating activation of the SPF and MPF activities of $p34^{cdc2}$. One hypothesis to explain this alternation of activities would be that G1 and mitotic cyclins are synthesized at different points in the cell cycle. But evidence that derivatives of mitotic cyclins can perform the function of G1 cyclins suggests there must be other differences between SPF and MPF (D. Lew and S. I. Reed, personal communication).

Broek *et al.* uncovered the role of *cdc2* in the alternation of mitosis and S phase by looking for mutants in the fission yeast, *Schizosaccharomyces pombe*, which would replicate their DNA twice without an intervening mitosis. Two such mutants were isolated, and both of them turned out to be previously known alleles of *cdc2*. Like all known alleles of *cdc2*, these mutants arrest most cells at G2 when an exponentially growing culture is shifted to the non-permissive temperature. Surprisingly, when these cells are returned to the permissive temperature, some of them did not immediately pass through mitosis but instead re-replicated their DNA and then passed through mitosis, yielding diploid cells. The fraction of cells that re-replicate their DNA can be increased by either starving the cells during the arrest or subjecting them to a brief heat pulse at 49 °C after arrest. Both of these treatments would be expected to increase the rate of protein degradation, implying that re-replication requires the physical destruction of $p34^{cdc2}$. Analysis of protein levels by

Broek *et al.* confirmed that the treatments that induce re-replication do indeed lead to the disappearance of $p34^{cdc2}$.

Taken together, these experiments suggest that the *cdc2* molecule itself carries a memory of its cell-cycle history (see figure). Immediately after mitosis the molecule is in a form dubbed the S form that can be converted into active SPF and induce DNA replication. At some point after Start, and before the completion of DNA replication, the $p34^{cdc2}$ molecule is converted into an M form that can be converted into active MPF but not into SPF. This reaction guarantees that mitosis will follow S phase. Passage through mitosis completes the cycle by converting the M form back to the S form. The diploidization of *cdc2* mutants can be explained by the destruction of the M form during heat shock, followed by



Models of the early embryonic and post-embryonic cell cycles. In the early embryonic cycle the alternation of mitosis and interphase is controlled solely by regulating the activity of maturation-promoting factor (MPF), and the block to re-replication of the DNA is removed as a result of the activation of MPF. In the post-embryonic cell cycle $p34^{cdc2}$ exists in two forms: the M form which can be converted into active MPF and the S form which can be converted into active Start-promoting factor (SPF). Active MPF stimulates the formation of the S form and active SPF stimulates the regular alternation of S phase and mitosis. Feedback controls regulate the activation and inactivation of MPF to ensure that mitosis does not begin until DNA replication is complete, and anaphase does not occur until assembly of the mitotic spindle is finished.

synthesis of new $p34^{cdc2}$ in the S form after the cells are returned to the permissive temperature. These experiments also imply that the default form of newly made $p34^{cdc2}$ is the S form, and that the presence of the M form is either dominant to the S form or induces the conversion of $p34^{cdc2}$ molecules from the S to the M form. From the observation by Broek *et al.* that the distinction between S and M forms is maintained even under conditions when $p34^{cdc2}$ is inactive and the cyclin associated with it has been degraded, it would seem that the two forms are distinguished by post-translational modifications on $p34^{cdc2}$.

The re-replication of *cdc2* mutants also sheds light on the 'block to re-replication' that prevents any segment of the genome from being replicated more than once in a given S phase⁵. Broek *et al.* show that the second round of replication of *cdc2* mutants, like the first, requires passage through Start, suggesting that passage through Start may be required to remove the block to re-replication. Because these cells have not passed through mitosis, passage through Start must be sufficient to remove the block to re-replication.

How is DNA replication regulated in the rapid cell cycles of early embryos? In these cell cycles DNA replication begins immediately after the exit from mitosis and there is no evidence for a discrete Start (although $p34^{cdc2}$ activity may be required very early in interphase⁶) and the block to re-replication is removed during mitosis⁷. I suggest that embryonic cycles represent a stripped-down (and perhaps evolutionarily old) cell cycle in which the activation of MPF directly removes the block to re-replication and $p34^{cdc2}$ is locked in the M form (see figure). By comparison, in somatic cells the removal of the block is no longer under the control of MPF, and the presence of the S form of $p34^{cdc2}$ and G1 cyclins allows Start to exist as an independent cell-cycle transition that controls the timing of replication.

Accurate reproduction of cells requires that DNA replication must be completed before mitosis begins. What is the relationship between this control and the regular alternation of SPF and MPF activity? Two possibilities exist. I favour a scheme in which the conversion of the S form to the M form of $p34^{cdc2}$ is an intrinsic property of the cell-cycle engine, is directly induced by the appearance of SPF, and is independent of the activation of DNA replication by SPF (illustrated in the figure). In this case the presence of un-

1. Broek, D., Bartlett, R., Crawford, K. & Nurse, P. *Nature* **349**, 388–393 (1991).
2. Murray, A.W. & Kirschner, M.W. *Science* **246**, 614–621 (1989).
3. Nurse, P. *Nature* **344**, 503–508 (1990).
4. Wittenberg, C., Sugimoto, K. & Reed, S.I. *Cell* **62**, 225–237 (1990).
5. Rao, P.N. & Johnson, R.T. *Nature* **225**, 159–164 (1970).
6. Blow, J.J. & Nurse, P. *Cell* **62**, 855–862 (1990).
7. Newport, J.W. & Kirschner, M.W. *Cell* **37**, 731–742 (1984).

replicated DNA must prevent the conversion of the M form to active MPF, rather than controlling the conversion of the S to the M form. The second possibility is that the conversion of the S to the M form cannot occur until DNA replication has been completed and is itself part of the system of checks ensuring replication is completed before mitosis begins.

Finally, there is the question of how some cells suppress the regular alternation of S and M phases. Cells become poly-

ploid by allowing many rounds of DNA replication without intervening mitoses. Do these cells act like the diploidizing mutants in *cdc2*, by degrading $p34^{cdc2}$, at the end of each round of DNA replication to destroy the M form, or do they have some other means for inactivating the block to re-replication? □

Andrew W. Murray is in the Department of Physiology, University of California, San Francisco, California 94143-0444, USA.

PARTICLE PHYSICS

Glueballs and exotic matter

Frank Close

If photons — the quanta of electromagnetic radiation — were electrically charged, positive and negatively charged photons could attract and form neutral clusters of light. But such 'photon-balls' do not exist as photons are electrically neutral. However, the quark theory of strongly interacting particles, such as neutrons and protons, implies that there should exist new forms of matter such as 'glueballs' similar in spirit to these imaginary photon-balls. This idea has been around for nearly 20 years and searches for glueballs have been going on almost as long. There are still no unambiguous sightings, yet — as contributions at recent meetings* showed — there is no slackening of interest in the idea.

The structure of bulk matter, molecules and individual atoms involves the underlying attractions and repulsion of positively charged atomic nuclei and their surrounding negatively charged electrons. Disturbing these electromagnetic forces liberates radiation which emerges as quantum bundles — massless photons. The relativistic quantum theory of the electromagnetic field is quantum electrodynamics or QED for short.

Our picture of the more powerful forces at work within the atomic nucleus has been developing for over 50 years. We know that the nucleus's constituent protons and neutrons are not fundamental, but are built from deeper constituents — the quarks. These are electrically charged but also carry a further charge known as colour. Whereas electrical charge comes in positive or negative forms, the amounts measured by simple numbers (a U(1) theory), the quarks' colour charge can have any of three varieties, described by a 3×3 unitary matrix (an SU(3) theory). To heighten the analogy with QED one can think of the quarks as having positive colour charge, and their antiparticle counterparts — the antiquarks — negative charge. The

resulting theory of colour-induced forces is called quantum chromodynamics (QCD) by analogy with QED.

The chromostatic forces — attractions between unlike colours — naturally draw quark to antiquark forming ephemeral states of 'quarkonium' (by analogy with positronium, the electron-positron 'atom'). The psi, upsilon and pion are examples of these states. The threefold colour charge gives further possible attractions among unlike colours. It is possible for three quarks to attract one another so long as each carries a different colour. Thus one naturally finds stable clusters of three quarks, of which the proton and neutron are the best known examples. This simple picture — quark plus antiquark or three-quark clusters — accommodates essentially all of the hundreds of short-lived sub-nuclear states discovered in half a century. Yet there is a tantalizing missing ingredient. This is where the glueballs come in.

QCD implies that there will be a colour radiation analogous to electromagnetic radiation; the quantum bundles are known as gluons (contrast the photons). So far it all seems similar to QED. But the presence of *three* colours for the quarks and the SU(3) ('non-Abelian') structure of QCD implies that the gluons carry colour themselves. Here we see an essentially new structure in QCD, richer in detail than that of QED. The photons of QED are uncharged and travel through space independently; the gluons of QCD are colour-charged and can mutually interact *en route*. This causes the long-range nature of the forces between quarks (transmitted by the exchange of gluons) to differ from the inverse-square law of QED — a satisfying feature that probably underwrites the close confinement of quarks. But it also implies that bound states of pure glue (the glueballs) can exist.

An obstacle for glueball hunters is that the overall properties of these states seem to be very similar to many conventional quarkonium states, so that their inner constitution is not easily manifested.

Making explicit predictions of their properties from QCD is currently beyond our analytical ability and so theorists have tended to construct tractable models that contain what are thought to be the important features of the dynamics. Although the various models tend to differ in detail, nonetheless they make predictions that agree in several broad features. The lightest glueballs are predicted to occur in the region of 1–2 GeV (one to two times the mass of a proton), to be short lived, and to have total intrinsic angular momentum (spin) of 0 or 2 units, where the proton's spin is half a unit. Their wavefunctions are eigenstates under (effectively unchanged by) spatial reflection (parity); the spin-0 with positive parity (0^+) is uniformly predicted to be the lightest; a spin-0 with negative parity (0^-) or the spin-2 positive parity (2^+) is expected to be the next heaviest.

For several years there have been three particular experimentally observed states with these quantum numbers that are *prima facie* candidates and fit the predictions but there has been much argument over how sure one can be of describing these as glueballs. Several recent developments add support to the belief that quarks alone are insufficient to describe all of particle spectroscopy. At the meeting *Hadron 90*, high-quality data from CERN and the US laboratories were announced that greatly elucidated quarkonium spectroscopy in the 1–2 GeV mass region; a result of this is that states are now being identified that do not appear to fit in with the simple quark picture.

Advances in simulating QCD theory on computers where space-time is treated as a collection of discrete points (lattice QCD) were reported at the Rutherford meeting (R. Kenway, Edinburgh University). These computations are now giving encouraging agreement with the masses of known states and predict that the lightest glueballs have 0^+ and 2^+ quantum numbers, the latter being some 50 per cent more massive than the former. These predictions are tantalizingly akin to data where 2^+ (1.7 GeV) and 0^+ (around 1 GeV) have been favourite candidates for gluonic states. However, the lattice calculations still suffer from severe approximations — the glueball states are computed with the assumption that no quarks exist.

This also relates to a problem for interpretation of data. It is unrealistic to imagine that physical states will be pure glue; quantum mechanics implies that there will probably be a mixing of character with quarkonium states whose quantum numbers are the same as those of the pure glue states. So the immediate strategy is to try to identify several states, with the same spin and parity quantum numbers, and hopefully find that they are too numerous to fit into a simple quarkonium spectroscopy. News at

* Rutherford—Appleton Laboratory Annual Review Meeting, 17–19 December 1990: *Hadron 90*, St Goar, Germany, 3–8 September 1990.

Hadron 90 is that this seems to be the case, at least for the 0^- quantum number ('pseudoscalar states'). This begins to add empirical encouragement to many theorists who have argued (the most recent in a long line being G. J. Gounaris and J. E. Paschalis; *Phys. Lett.* **B251**, 634–638; 1990) that detailed structures in spectroscopic data around 1.4 GeV mass are best understood if there is a 0^- glueball.

The importance in identifying a detailed spectroscopy of glueballs is that sighting bound states of gauge fields (such as glueballs), as distinct from those of fermions (the conventional 'material' particles), would be entirely new. We have good evidence that the underlying mathematical structure of QCD and QED are very similar but the richness of the SU(3) non-Abelian structure of QCD

implies that there are novel possibilities such as non-trivial vacua, instantons and probably further profound consequences that will be revealed only through experiment. To identify these it is important to isolate the spectroscopy of the glueballs. Theory suggests that glueballs are accessible at rather low energies, well within the reach of existing laboratories. It will be the advances in detector technology and particle beam quality and intensity that will be the keys to opening up this area of particle physics, and will offer the hope of deepening insights into the physics of non-Abelian gauge field theories. □

Frank Close is at the Rutherford–Appleton Laboratories, Chilton, Didcot OX11 0QH, UK, and Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, Tennessee 37831, USA.

VACCINE DEVELOPMENT

Real and imagined dangers

Gordon Ada

EARLIER this month a report by Oehen *et al.* appeared in *Science*¹, under the title "Vaccination for Disease", in which the authors describe some experiments with lymphocytic choriomeningitis virus (LCMV) in mice. This virus contains two main protein types, glycoproteins and a nucleoprotein. In some experiments, pre-immunization of mice with vaccinia virus–LCMV single protein constructs exacerbated the disease that occurred when the mice were challenged with LCMV administered intracerebrally. In contrast, pre-immunization with intact LCMV protected the mice. Although these experiments are valid and helpful lessons can be drawn from them, the very title of the paper might be inappropriately taken as a cause for alarm, particularly as chimaeric viruses (or bacteria) have already been successfully used for vaccination.

Elsewhere, I have proposed² that a vaccine should meet four immunological requirements, three of which are relevant here. One is that the vaccine should contain a sufficient number of different T-cell epitopes that T-cell responses will be achieved in all members of a genetically diverse outbred population. Inclusion of several different proteins in the preparation increases the number of available epitopes. A second requirement is that the vaccine should generate a large pool of memory T (and B) cells; the greater the size of this T-cell pool, the quicker the appearance of effector T-cell activity generated from the memory cells by the challenge virus. The results of Oehen *et al.* are consistent with these points, as shown in their Figs 1 and 2C. As expected, intact LCMV which contains all the viral antigens was more protective than either of the chimaeric vaccinia virus constructs.

A third requirement of a vaccine is the generation and persistence of high titres of neutralizing antibody so that the great bulk of a challenge virus, at any time after immunization, is prevented from infecting susceptible cells. Oehen *et al.* did not measure such antibody generated by any of the immunization schedules, but it is very likely that immunization with whole virus would have achieved this most effectively. This, as well as the rapid appearance of effector T cells after challenge (their Fig. 2C), may have been a deciding factor in preventing the deaths of animals that had been preimmunized with intact virus.

Several, more general, points should also be made. First, it is inevitable that T-cell activation will result in some immunopathological effects. Interleukins, which are defined as mediators of inflammation, are secreted by both CD4⁺ and CD8⁺ cells. The trick is to limit the effect as far as possible and this can be done by minimizing the extent of infection that occurs upon challenge. The risk of immunopathology becomes greater when, for reasons such as great antigenic variation, the virus largely escapes the neutralization of infectivity by preformed antibody.

Second, LCMV infection is an unusual model. Challenge by intracerebral injection, used in the experiments of Oehen *et al.*, is an unnatural route of infection. Virus inoculated into the brain can form a reservoir not easily available to the immune system. Other than vector-borne agents, most agents infect through a mucosal surface and there is increasing interest in developing vaccines which will stimulate a secretory IgA response as this is the first line of defence.

Third, there are many examples with

other viruses where transfer of virus-specific cytotoxic T cells before or after administration of a lethal dose of virus has cleared the infection and prevented death. In addition, there are many examples in model systems where immunization with a chimaeric virus preparation has provided protection against a challenge dose of the infectious agent which was the source of the foreign antigen(s)³. Two outstanding examples are vaccinia virus constructs which contain antigen from rabies virus^{4,5} or from rinderpest virus⁶ — foxes immunized with the vaccinia–rabies virus construct survived challenge with many thousands of lethal doses of rabies virus; cattle immunized with the vaccinia–rinderpest virus construct were protected against challenge by a thousand lethal doses of rinderpest virus.

Most vaccines currently in use are to control acute infections. Among the most successful are those consisting of live, attenuated viruses which induce cytotoxic T-cell responses. Many of the infectious agents that are the target for vaccine development cause chronic persisting infections. The observations of Oehen *et al.* may be relevant in some of these cases, for example in HIV infections where there is great antigenic variation in the envelope protein. A candidate vaccine now in phase 1 human trials is a construct composed of vaccinia virus and the HIV envelope protein. The rationale behind this approach is to generate cytotoxic T-cell responses which would limit infection caused by either free virus or infected cells. It remains highly desirable that such a vaccine should still aim to induce cross-protective neutralizing antibodies to limit the extent of viral infection upon initial exposure, and this may now be possible⁷. The current development of a vaccinia virus construct containing three HIV antigens is a useful further approach to achieve rapid elimination of infected cells.

Although it may never be possible to guarantee that a given approach will always result in a completely safe vaccine, enough is now known about the important immunological requirements for vaccines, especially for viral infections, to proceed with confidence in developing the next generation. The use of live vectors, viruses or bacteria, is one of several approaches in use; it may turn out to be the best in some situations. □

Gordon Ada is in the School of Hygiene and Public Health, The Johns Hopkins University, Baltimore, Maryland 21205, USA.

- Oehen, S., Hengartner, H. & Zinkernagel, R.M. *Science* **251**, 195–198 (1991).
- Ada, G.L. *Lancet* **335**, 523–526 (1990).
- Flexner, C. & Moss, B. in *New Generation Vaccines* (eds Woodrow, G.C. & Levine, M.M.) 189–206 (Dekker, New York, 1990).
- Blancou, F. *et al.* *Nature* **332**, 373–375 (1986).
- Brochier, B. *et al.* *Vaccine* **8**, 101–104 (1989).
- Yilma, T.D. *et al.* *Science* **242**, 1058–1061 (1988).
- Javaherian, K. *et al.* *Science* **250**, 1590–1593 (1990).

Physarum — C the difference?

Brenda L. Bass

In recent years there have been many reports of an unprecedented way to alter gene expression at the RNA level (reviewed in refs 1 and 2). The phenomenon, termed RNA editing, results in the insertion, deletion or substitution of nucleotides in particular RNA molecules, and thus has the remarkable ability to alter, or even create, specific codons. In its most impressive form, RNA editing can create a translatable messenger RNA from the transcript of a gene that completely lacks a functional open reading frame. On page 434 of this issue³, Mahendran *et al.* describe such a case of RNA editing, but one with a difference. The authors find that in the slime mould *Physarum polycephalum* the mitochondrial mRNA encoding the α subunit of ATP synthase is edited at 54 sites by the

tically from that seen in mammalian apolipoprotein B mRNA^{4,5} and certain plant mitochondrial mRNAs^{6,7}. The non-templated uridines in these other RNAs derive from templated cytidine residues, presumably by deamination. Certain viral RNAs have inserted guanines⁸, but, because these nucleotides always follow a stretch of guanines, they are thought to arise cotranscriptionally, as a result of polymerase stuttering. Although only proven directly for the edited apolipoprotein B mRNA^{9,10}, all other types of RNA editing are thought to occur post-transcriptionally. The sequences surrounding the editing sites in the ATP synthase mRNA are not characterized by stretches of cytidines, but of course it remains to be seen if this editing occurs post-transcriptionally.

and at the third position of the codon.

The pattern of editing observed in the *Physarum* RNA is suggestive, but of what is unclear. As for all examples of RNA editing, the crucial question concerns the source of the information that dictates the precise editing seen in the final product. Is the information contained in the pre-edited transcript itself, or is it provided by a separate unconventional template? In kinetoplasts it seems that editing information is provided by small, mitochondrion-encoded RNAs called guide RNAs (gRNAs; see a recent News and Views article¹¹). If G·U as well as normal Watson-Crick base-pairs are allowed, gRNAs turn out to be complementary to short regions of fully edited transcripts. Several models have been proposed whereby gRNAs, together with an endonuclease, a uridylyl transferase and an RNA ligase, direct the production of fully edited RNAs. Mahendran *et al.*³ state they are searching for antisense RNAs that might be analogous to the gRNAs of kinetoplasts. Of course, whereas the gRNAs of kinetoplasts seemingly use two nucleotides, G and A, to direct the insertion of uridines, the putative gRNAs of *Physarum* would be limited to the use of a single nucleotide, G, as a template for cytidines. In addition, a cytidylyl transferase rather than a uridylyl transferase would be required for the process.

Editing in *Physarum* could certainly involve gRNAs, but it is premature to rule out the possibility of information in the pre-edited RNA being involved. In fact, the nonrandom characteristics of *Physarum* editing, such as the spacing and preference for a 5' purine-pyrimidine dinucleotide, may be indicative of a so far unrecognizable primary or secondary structure that serves to mark sites for editing. Existing studies indicate that coding capacity as well as secondary structure can place a selectional constraint on mRNA sequence¹², and it has been noted that the third position of the reading frame could serve to fine tune RNA structure with a minimal disruption of coding capacity. Many organisms apparently produce the mature mRNA for the α subunit of ATP synthase without editing,

UUAUGUGAUUAUGGUUUUGUUUUUUAUUGGUAUUUUUUAGAUUUA
Δ

GUCAAUCGGUCAAUAUUUCUGUCAAAGAUGGUGUUGCUUUUGUACAGGACUU

Comparison of RNA editing in *Trypanosoma brucei* (top, mRNA for cytochrome c oxidase III mRNA) and *Physarum polycephalum* (bottom, mRNA for α subunit of ATP synthase). Bold letters indicate inserted residues; triangle indicates a uridine deletion.

insertion of cytidine residues. This is the first observation of RNA editing in *P. polycephalum*, but more significantly, it is the first case of RNA editing that produces non-encoded cytidine residues.

The work began with the identification of a complementary DNA (cDNA) that was homologous to a single site in the *Physarum* mitochondrial DNA. The complete cDNA sequence contained a single reading frame of 528 codons whose deduced amino-acid sequence showed that it coded for the α subunit of ATP synthase. In contrast, although the mitochondrial DNA sequence was clearly homologous, the region corresponding to the open reading frame was characterized by many -1 frameshifts. Such discrepancies between the sequence of a cDNA and the corresponding genomic sequence are the classic hallmark of RNA editing; thus, after convincingly demonstrating that the RNA was not encoded on a conventional DNA template located outside the mitochondrion, the authors concluded that they had discovered yet another cryptogene.

How does the editing observed in *Physarum* fit in with previous examples? First, it clearly distinguishes itself as the only known example of editing that results in non-encoded cytidine residues. Further, because the cytidine residues seem to arise by insertion, rather than modification of a templated base, the editing in *Physarum* must differ mechanis-

By a process of elimination it is tempting to conclude that editing in *Physarum* is similar to the insertion and deletion of uridine residues found in many mitochondrial mRNAs of kinetoplastid protozoa such as *Trypanosoma brucei*. The differences, however, are striking. The figure shows a small region of the extensively edited cytochrome c oxidase III mRNA of *T. brucei*¹¹ and a region of the *P. polycephalum* ATP synthase mRNA that includes the first three editing sites. The large bold letters indicate insertions, and the triangle indicates a uridine deletion. Compared with the almost chaotic insertion and deletion of uridine residues observed in the edited RNAs of kinetoplasts, the edited RNA of *Physarum* suggests a calm order. Whereas multiple uridines can be inserted at each editing site in the kinetoplast RNAs, only a single cytidine is inserted at each of the 54 editing sites of the ATP synthase mRNA. The *Physarum* mRNA shows no signs of deletions. In kinetoplast RNAs, editing often occurs at each phosphodiester bond for several contiguous residues; in contrast, the editing sites in the ATP synthase mRNA are spaced no closer than 11 nucleotides and no farther than 54 nucleotides (the average is 26 nucleotides). The authors further emphasize the relatively nonrandom nature of *Physarum* editing by noting that at most of the editing sites the cytidines are inserted 3' to a purine-pyrimidine dinucleotide

1. Scott, J. *Curr. Op. Cell Biol.* **1**, 1141-1147 (1989).
2. Benne, R. *Trends Genet.* **6**, 177-181 (1990).
3. Mahendran, R., Spottswood, M.R. & Miller, D.L. *Nature* **349**, 434-438 (1991).
4. Powell, L.M. *et al. Cell* **50**, 831-840 (1987).
5. Chen, S.-H. *et al. Science* **238**, 363-366 (1987).
6. Gualberto, J.M., Lamattina, L., Bonnard, G., Weil, J.H. & Grienberger, J.M. *Nature* **341**, 660-662 (1989).
7. Covello, P.S. & Gray, M.W. *Nature* **341**, 662-666 (1989).
8. Cattaneo, R., Kaellin, K., Bacsko, K. & Billeter, M.A. *Cell* **56**, 759-764 (1989).
9. Driscoll, D.M., Wynne, J.K., Wallis, S.C. & Scott, J. *Cell* **58**, 519-525 (1989).
10. Chen, S.H., Li, X.X., Liao, W.S., Wu, J.H. & Chan, L. *J. Biol. Chem.* **265**, 6811-6816 (1990).
11. Feagin, J.E., Abraham, J.M. & Stuart, K. *Cell* **53**, 413-422 (1988).
12. Blackburn, E.H. *Nature* **346**, 609-610 (1990).
13. Knings, D.A.M., van Duijn, L.P., Voorma, H.O. & Hogeweg, P.J. *theor. Biol.* **127**, 63-78 (1987).

so it may be interesting to compare the sequence of these RNAs, in particular the wobble positions, with the sequence of the *Physarum* transcript. Differences between the RNAs may reflect selectional constraints placed on the *Physarum* transcript by primary or secondary structures involved in the editing. Of course, a serious search for editing information in the pre-edited *Physarum* RNA would be

complicated. However, given the rapid rate of progress in our understanding of RNA editing, it seems probable that the necessity of undertaking such a search will soon become evident. □

Brenda L. Bass is in the Department of Biochemistry, University of Utah, 50 North Medical Drive, Salt Lake City, Utah 84132, USA.

LOMA PRIETA EARTHQUAKE

Putting the pieces together

William L. Ellsworth

WHEN it rocked the San Francisco Bay area on 17 October 1989, the Loma Prieta earthquake (of magnitude, M_w , 6.9) was embraced as fulfilling a long-term forecast for the rupture of the San Andreas fault in the southern Santa Cruz mountains¹. Now, after an additional year of digging into old geodetic data — and, quite literally, into the fault itself — a new picture is beginning to emerge of comparatively infrequent rupture of a secondary fault trace along this reach of the San Andreas. These developments sparked a lively debate about the significance of this event and the prospects for additional activity along this segment of the fault at a recent meeting of the American Geophysical Union*.

The Loma Prieta earthquake ruptured a 40-km-long segment of the San Andreas fault along the southernmost end of the great 1906 earthquake ($M_w = 7.7$) rupture, and was the first significant event to strike along it in the past 83 years. Faulting in 1989 propagated upward along a dipping plane to intersect the San Andreas at about 10 km depth. Right oblique reverse faulting on this plane raised the southwest (Pacific Plate) side of the fault 1.1 m and translated it 1.6 m to the northwest, relative to the northeast (North American Plate) side (see box).

The occurrence of an earthquake along this segment of the fault had been predicted by a number of workers during the previous decade, principally on the basis of the time it would take to recover the strain that had been released there in 1906².

Early on, it was speculated that the actual earthquake was an unusual event, as the large proportion of vertical slip did not fit the conventional model of transform fault motion along the San Andreas fault³. Repetition of this event at a rate needed to match the local plate velocity should build high mountains, not the modest southern Santa Cruz mountains which rise to only 1,100 m.

Part of the mystery about the behaviour of this segment of the fault stems from

uncertainties about its behaviour in the 1906 earthquake. Although surface faulting was clear elsewhere along the fault⁴, the reported effects in the southern Santa Cruz mountains were dominated by secondary ground failures, many now suspected of being a gravitationally driven response to the shaking, just as occurred in the 1989 event. Thus, we have little direct evidence of how plate motion is accommodated locally.

With surface faulting also absent in 1989, Paul Segall and Mike Lisowski⁵ turned to the analysis of geodetic deformations induced by each event. They found that the two earthquakes displaced the nearest mark, Loma Prieta peak, in different directions, thereby demonstrating that the two events are not simply scaled versions of a single, characteristic displacement. The horizontal displacement in 1906 was nearly parallel to the fault trace, whereas the 1989 earthquake clearly moved the mark towards the fault, as well as parallel to it.

This oblique motion is well understood in terms of the mechanism of the Loma Prieta earthquake. At a minimum, 1906

must have involved 2.5 ± 0.4 m of slip parallel to the fault at shallow depth (0–10 km), whereas little slip occurred above 4–5 km in 1989. It is possible, however, that the 1989 fault could have moved in the same manner in 1906 as well, as its 0.1 m transverse displacement of Loma Prieta peak would be undetectable in the 1906 measurements.

Grant Marshall (US Geological Survey, Menlo Park) strengthened the case that something was unusual about the 1989 event when he presented the elevation changes determined by levelling. He corroborated the speculations (ref. 6; G. Valensise, Istituto Nazionale di Geofisica, Rome; S. N. Ward, University of California, Santa Cruz) that Loma-Prieta-type events uplift the marine terraces along the Santa Cruz coastline. As Valensise noted, the terraces can be fully built by Loma Prieta events occurring once every 600 years, assuming no other displacements add to or subtract from the measured uplift. Valensise also reported on preliminary investigations of the tectonic uplift of the youngest marine rocks in the area, the 5-million-year-old Purisima Formation. His results suggest that uplift proceeded at a rate of 0.5 mm yr^{-1} over the past 3 million years, also indicating the infrequent occurrence of Loma-Prieta-type events.

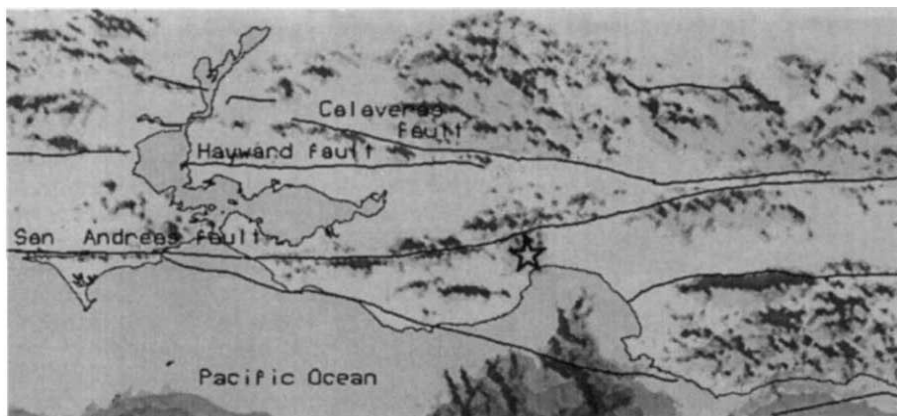
If recurring Loma Prieta events account for only 10–20 per cent of the expected San Andreas motion, where is the rest of it? New trenching investigations in the southern Santa Cruz mountains by David Schwartz and co-workers (US Geological Survey, Menlo Park) have encountered probable 1906 surface faulting at one site. If palaeoseismic investigations planned for the spring prove successful, we stand to learn if the 1906 event completes the picture, or if other hidden faults lies within

Transform faults and vertical tectonics

TRANSFORM faults accommodate lateral translation between two rigid lithospheric plates without either convergence or divergence along the plate boundary. Strictly speaking, this is true only when the transform fault describes a small circle about the Euler pole of relative motion between the two plates. When a transform fault deviates from the small circle azimuth, parts of the colliding plates must be removed if the deviation places the boundary into compression. Conversely, a hole must be filled if the plates diverge along the canted transform fault. The rate of destruction or void-filling is equal to the slip rate along the properly oriented transform multiplied both by the width of the step, as measured normal to the plate boundary, and by the depth of the fault.

Along most of its 1,200-km length, the San Andreas fault lies within a few degrees of small circles (see figure overleaf). But, there are important exceptions, such as the 'Big Bend', where the fault strike deviates in either a clockwise or anticlockwise manner, placing the plate boundary into extension or compression, respectively⁶. The step-over along the fault in the southern Santa Cruz mountains measures about 10 km. Here, the excess crustal volume attempting to pass through the bend amounts to about $3,000 \text{ km}^3$ per million years, and is locally balanced, at least approximately, by the 'hole' formed along the Calaveras fault, at the bifurcation of the single fault strand into the parallel Hayward and San Andreas faults. The reverse motion in the Loma Prieta earthquake removed about a century's worth of the excess crust by contributing to the construction of these young mountains. □

*American Geophysical Union Fall Meeting, San Francisco, 3–7 December 1990.



Shaded-relief map of central California and Pacific Ocean floor shown in oblique Mercator projection about the Euler pole of motion between the North American and Pacific Plates (49.6° N, 76.7° W). Throughout much of the region the plate boundary follows small circles (lines of latitude) about the Euler pole, principally as the San Andreas and Hayward faults. The Loma Prieta earthquake (star) occurred in the southern Santa Cruz Mountains, where the 10° anticlockwise bend of the San Andreas fault places the plate boundary into compression. Note the low topography along the symmetrically branching Calaveras fault, where plate motion creates a deep basin.

the southern Santa Cruz mountains.

But what of the earthquake forecast? Several voices have been raised suggesting that claims of success are invalid, because the Loma Prieta earthquake occurred sooner than anticipated or was not directly on the main trace of the fault. It is fair to say that none of the published forecasts called this event correctly in all respects, and none foresaw the reverse component of motion. As Wayne Thatcher (US Geological Survey, Menlo Park) has emphasized, the forecasts with the most accurate timing critically depend on the questionable, in his view, adoption of the displacement (1.5 m) measured at Wrights Tunnel rather than the larger geodetically averaged displacement as the appropriate value for the 1906 earthquake.

We shall never know what slip, if any, occurred on the Loma Prieta fault in 1906, nor can we be guaranteed that the spacing of such events is regular in time. Thus, in retrospect, we are fundamentally limited in our ability to tie these two events together quantitatively.

Forecasting the future, especially with little in the way of either theory or data, is an unsure proposition, at best. To me, the Loma Prieta event marks an important success for earthquake prediction, albeit a qualified one, as the evidence suggests that between two-thirds and all of the strain released in 1906 had reaccumulated by the autumn of 1989 — more along this section than anywhere else along the 1906 rupture. The complexity of both the fault zone and the forces that drive it, and our limited knowledge of each, all but guarantee uncertainty in predicting when and where dynamic slip will begin or end along the fault. Thus, we should not be terribly surprised if many of the details — including the precise slip plane — vary from cycle to cycle. Although we may hope that simple concepts, such as “time predictability” grossly characterize the

process, it is not proved. Indeed, the horizontal slip in 1989 released all of the strain expected to have accumulated since 1906, nominally making this component of displacement “slip predictable”.

Loma Prieta poses another and perhaps ultimately more important dilemma for the would-be forecaster, as the elastic strain accumulated in the upper 4–5 km of the crust since 1906 was not released in October 1989, and there may still be significant potential for another strong event on this part of the San Andreas fault¹. This possibility was considered by the Working Group on California Earthquake Probabilities², who concluded that a magnitude 7 event was unlikely (probability less than 0.01) and a magnitude 6–6½ event was considered possible, though a formal probability was not assigned. Others, including J. Behr (CIRES, University of Colorado) and co-workers³ have reached a similar conclusion.

Some small comfort to the residents of the southern Santa Cruz mountains may be found in the observation that large magnitude earthquakes along the San Andreas fault system have historically involved energy release at depths well below 5 km. It is just in this deeper zone where the Loma Prieta earthquake released the stored strain along the fault zone. □

William L. Ellsworth is at the US Geological Survey, Menlo Park, California 94025, USA.

1. US Geological Survey *Science* **247**, 286–293 (1990).
2. Working Group on California Earthquake Probabilities, *US Geol. Sur. Circ.* 1053 (1990).
3. Kanamori, H. & Satake, K. *Geophys. Res. Lett.* **17**, 1179–1182 (1990).
4. Lawson, A.C. et al. *The California Earthquake of April 18, 1906* (Carnegie Inst., Washington, 1908).
5. Segall, P. & Lisowski, M. *Science* **250**, 1241–1244 (1990).
6. Anderson, R.S. *Science* **249**, 397–401 (1990).
7. Shimazaki, K. & Nakata, T. *Geophys. Res. Lett.* **7**, 279–282 (1980).
8. Behr, J., Bilham, R., Bodin, P., Burford, R.O. & Bürgmann, R. *Geophys. Res. Lett.* **17**, 1445–1448 (1990).
9. Bilham, R. & King, G. *Geophys. Res.* **94**, 10204–10216 (1989).

Read my lips

ESPIONAGE agents, it is said, can monitor a conversation in a house from a point outside it by bouncing a laser beam off a window. The conversation vibrates the window like a microphone diaphragm; the Doppler modulation thus imposed on the reflected beam can be read from some way away. Daedalus is now using this cunning principle in a novel hearing aid. Conventional hearing aids are very imperfect. They disable the user's normal ‘cocktail-party effect’ discrimination; the jumbled mixture of amplified sounds which they deliver is very hard to interpret.

By contrast, Daedalus's laser-based hearing aid ignores sound completely, and responds only to vibrating surfaces. Its low-power solid-state infrared laser, mounted on a spectacle frame, projects a tight beam into the central region of the wearer's field of view: a speaker facing him, say. The speaker's mouth, and indeed his entire face, must vibrate in sympathy with the sound being generated inside it, and will modulate the laser light reflected from it. A photodiode receiver on the spectacle frame demodulates the returning signal, and drives a conventional amplifier and earpiece.

This wonderful device gives the wearer amazingly sensitive and directional hearing. Its range is almost unlimited; the laser beam is attenuated only slightly by distance. From the back of the most unruly lecture hall, it will reach out to the lecturer alone, and read his lips perfectly without interference. In the noisiest party the wearer will hear only the companion facing him. To hear anyone else, even across a roomful of yelling people, he need only turn and look. He may even concentrate his gaze on the damnable background music that is forcing everyone to shout in the first place. Vibrating from its emitted sound, the loudspeaker will return a clear and uncluttered signal. Hearing will be almost as well-resolved as sight.

Engineers and musicians, even those with normal hearing, will rush to buy the new deaf aid. How useful for a choirmaster or conductor to be able to concentrate on just one singer or instrument in the vast polyphonic array before him! How convenient for a motor mechanic or service engineer to scan his gaze over the throbbing complexity of some malfunctioning monster, and hear the noise emitted by each component in turn! A more advanced, telescopic version would be ideal for diagnosing big industrial installations without leaving your office, or space rockets on the launch pad that cannot be safely approached. Even commanders in the field could use it to read the sonic signatures of the military hardware facing them, and the orders being given to deploy it . . . but now we're back to espionage again.

David Jones

A bigger can of worms?

SIR—Maddox¹, commenting on a review by Joseph and Preziosi², suggested that there are reasons why Fourier's law should be regarded "as the crudest approximation to the truth". The assertion that a fundamental law is, or could be, flawed would be expected to lead to a heated debate; the absence of a response in the year since the articles were published suggests that whatever was troublesome was either so trivial that it has gone away, or so profound as to deter anyone from grasping the nettle. We have considered the problem of heat conductance and some of the wider theoretical implications, particularly as they affect biologists' thinking about 'diffusion'.

Although Maddox points out that "mathematically, Fourier's law . . . is strictly equivalent to the diffusion equation", Fourier's work on the problem of heat (energy) movement in metals³ antedates by many years Fick's theory of diffusion. Fick transposed Fourier's idea to the movement of solutes in liquids⁴ without apparently providing experimental evidence to justify this extrapolation.

Fick's law subsequently seems to have been elevated to the level of orthodoxy: more recent attempts to validate it have not been successful. As Robinson and Stokes remarked⁵, bringing a lamp up to an apparatus in which the object was to measure diffusion in an undisturbed solution in which all bulk flow had (they hoped) been eliminated over many days immediately introduces convection currents that would nullify their efforts.

Joseph and Preziosi criticized² Fourier's theory of heat propagation in solids because they believed he had overlooked or neglected a factor for the movement of heat by radiation. Fourier had not in fact overlooked radiant heat¹, but considered it a minor phenomenon in the circumstances and kinds of solids he was investigating. Article 433 of his theory states: "luminous heat penetrates solids and liquids and is gradually absorbed by them after traversing a certain distance . . . When this distance has finite value, the differential equations take a different form, which could offer no useful application unless based on experimental knowledge which we have not yet acquired."

That light passes through glass (solid) and emerges as light and radiant heat seems common-sense confirmation of the concept that wave energy moves through and heats solid matter. Moreover, Fourier said that the coefficients used in his theory as constants were in fact constant only within narrow ranges. Having so fastidiously constrained his theory, Fourier would probably agree that, used outside those limits, it might be "the crudest

approximation to the truth".

If, therefore, Fourier's mathematics are still considered inadequate in describing heat movement in metal bars, then Fick's transformation of it for material movement in liquid must also be suspect. We are therefore surprised that no one has drawn attention to the possibility that the diffusion equation may be an even bigger can of worms.

Biologists lay great emphasis on diffusion as the means whereby molecules move from one part of a cell or tissue to another. But diffusion should surely be seen as occurring in a field of force. It is too often invoked as a satisfactory explanation for molecular movement within cells and organisms, without experimental evidence to support it.

We wonder whether too much credence is placed on diffusion by biologists, whether or not diffusion theory is flawed. Diffusion in its simple (unassisted) form often has little relevance or place in meta-

bolic processes and offers very limited scope for their sensitive regulation. If Fourier can be mistakenly criticized for ignoring radiant heat, Fick can be called to task not only for his extrapolations, but for ignoring convection. At the molecular level, continuing faith in diffusion may be a legacy of traditional biochemistry, in which enzymatic reactions were seen as events occurring in homogenates or solutions where the law of mass action could be simply applied.

P. C. MALONE

273 Kingsbury Road, Erdington,
Birmingham B24 8RD, UK

D. N. WHEATLEY

Department of Pathology,
University Medical Buildings,
Foresterhill, Aberdeen AB9 2ZD, UK

1. Maddox, J. *Nature* **338**, 373 (1989).
2. Joseph, D. D. & Preziosi, L. *Rev. Mod. Phys.* **61**, 41–73 (1989).
3. Fourier, J. B. *Theorie Analytique de la Chaleur* (Oeuvres, Paris, 1822).
4. Fick, A. *Ann Phys. Lpz.* **170**, 59–86 (1855).
5. Robinson, R. A. & Stokes, R. H. *Electrolyte Solutions* (Butterworths, London, 1965).

N₂O production in the ocean

SIR—Recent measurements of the ¹⁵N:¹⁴N ratio ($\delta^{15}\text{N}$) of dissolved nitrous oxide (N₂O) in sea water have revived the controversy concerning the mechanisms of N₂O production in the ocean. The observed high $\delta^{15}\text{N}$ of N₂O has been interpreted differently by Yoshida *et al.*¹ and by Kim and Craig² to support its production through denitrification and nitrification, respectively. I propose an alternate mechanism involving the possible coupling of the two processes through nitric oxide (NO) at low oxygen concentrations which may as well explain the isotope data.

The issue of whether denitrification could be a significant mechanism of N₂O production has been addressed^{2,3}. Although the lack of information on the isotope composition of ammonia in the oxygen-minimum zone represents a serious shortcoming in the existing data sets, it seems fairly obvious that sizeable N₂O production should occur through an intermediate species which is enriched in ¹⁵N (and ¹⁸O). It is probable that hydroxylamine (NH₂OH), produced through nitrification, is one such compound, as suggested by Kim and Craig². However, instead of the direct production of N₂O through the oxidation of NH₂OH, there could be another intermediate, NO, which could be reduced to N₂O within the micro-environments. There are several points that support this pathway (NH₄⁺ → NH₂OH → NO → N₂O). (1) NO is believed to be produced during nitrification at low oxygen concentrations. Ward and Zafiriou

estimated⁴ that the vertically integrated NO flux could be as much as 13% of the nitrification flux in the eastern tropical North Pacific. (2) The maxima in NO concentration and turnover rate are located at about the same level as those of N₂O, near the boundaries of the oxygen-deficient layer⁴. This implies a close relationship between NO and N₂O. (3) NO produced during nitrification is significantly enriched (by up to 20‰ relative to NH₄⁺) in ¹⁵N, whereas N₂O produced during this process is severely depleted⁵. (4) The mechanism favoured by Kim and Craig cannot easily account for a large increase in the extent of N₂O production at low oxygen concentrations. This requires a reduction step: either 'nitrifier denitrification' (NO₂⁻ → N₂O) or the mechanism proposed here. The fact that the former process leads to large ¹⁵N depletion⁵ suggests that it may not be dominant.

The nitrification–denitrification couple can also account for the low yields of N₂O during the incubation experiments of Yoshida *et al.*¹. This could merely reflect the scarcity of large particles within which conditions favourable for NO reduction could develop.

The maximum in $\delta^{15}\text{N}$ is located significantly below the oxygen minimum². And as the maximal N₂O concentrations are found within the oxygen minimum¹, it is clear that the mechanism(s) leading to the greater ¹⁵N enrichment in N₂O in deep water may not also be mainly responsible for its production. It seems to me that there may be several mechanisms of N₂O production. As refs 1 and 2 both reveal lower $\delta^{15}\text{N}$ in N₂O in the upper layers compared with the deep sea, the pathway NH₄⁺ → NO₂⁻ → N₂O (ref. 5) might con-

tribute significantly to the N_2O flux in the upper few hundred metres. Within the oxygen-minimum zones, where most of the oceanic N_2O is concentrated⁶, the mechanism I propose might be dominant. Significant production might also occur through the mode favoured by Kim and Craig. However, the bulk of oceanic N_2O production could be through the nitrification–denitrification couple I propose.

S.W.A. NAQVI

National Institute of Oceanography,
Dona Paula, Goa 403 004, India

1. Yoshida, N. *et al.* *Nature* **342**, 895–897 (1989).
2. Kim, K.-R. & Craig, H. *Nature* **347**, 58–61 (1990).
3. Naqvi, S.W.A. & Noronha, R.J. *Deep-Sea Res.* (in the press).
4. Ward, B.B. & Zafiriou, O.C. *Deep-Sea Res.* **35**, 1127–1142 (1988).
5. Yoshida, N. *Nature* **335**, 528–529 (1988).
6. Codispoti, L.A. & Christensen, J.P. *Mar. Chem.* **16**, 277–300 (1985).

Ancient Arctic hunters

SIR—In 1989, a joint expedition from the Leningrad Institute of Archaeology and the Arctic and Antarctic Research Institute explored the ancient site on the Zhokhov isle (New Siberia isles, 76° N; Fig. 1). The existence of this remote site was known, but no details as to its dates

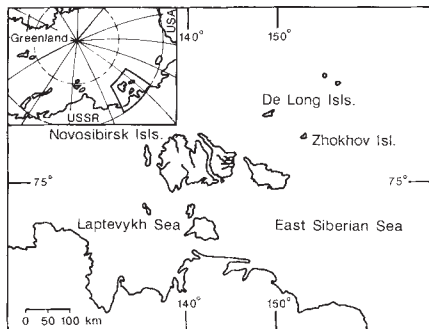


Fig. 1 Map of the Zhokhov isle region.

and type of occupation have previously been reported. Our data suggest that this area was occupied far earlier than has previously been thought.

The site we studied is in the southwest of the island, near the foot of a hill, which was probably used as an observation post or as a defence from the extreme weather. We found the remains of at least 13 dwellings within the site. The buildings were made of driftwood and covered by moss or sod. The walls were probably strengthened by vertical pieces of driftwood.

We excavated material from under one



Fig. 2. Bone implements found on Zhokhov isle. Double-edged (1, 2) and one-sided (3–5) tools; bone points (6–8) and fragment of polished mammoth ivory specimen (9). Scale bar, 3 cm.

dwelling which provides information about the culture of the people who lived in the site. Various material was used for the making of stone tools. Different kinds of flint, silicated schiststone, halcedone and obsidian were used: undoubtedly, obsidian and scanty flint of high quality were imported. The techniques of processing the stone is typical of the mesolithic period.

We collected a series of massive mammoth-ivory and antler hoe-shaped tools, and found specimens of hunting implements: needle-shaped arrow points and fragments of large inset tools — both one-sided and double-edged (Fig. 2).

The fauna remains were examined by A. Kasparov at the Leningrad Institute of Archaeology. Most are from reindeer (50%) and white bear (45%). The rest of the bones we could define are from sea-mammals (such as walrus and seal), wolf and birds. Thus, there were two main species of prey. This specialisation of hunting is unusual.

We dated the site (processing of specimens was carried out by Y. Swejentsev at the Leningrad Institute of Archaeology); most wood and bone specimens examined define the age of the site as 7,000 to 8,000 years before present.

Thus, the finds on the Zhokhov isle are changing our notions about the time and extent of the mastering of the high Arctic by man. Our results suggest that the first migration to these territories happened even earlier.

VLADIMIR PITUL'KO

Leningrad Institute of Archaeology,
Dvortsovaya nab. 18,
Leningrad 191065, USSR

VJATCHESLAV MAKEYEV

Arctic and Antarctic Research Institute,
Ulitsa Beringa 38, Leningrad, USSR

Provenance of HIV strains

SIR—Since your News report¹ on the experience of several laboratories, including mine, of accidental transmission of HIV in cell cultures, I have received some queries about whether any of our HIV isolates cannot be clearly distinguished from previously described strains. Of the viruses we have reported or distributed to other laboratories, CBL4 (HIV-1) and CBL20-24 (HIV-2) clearly differ from all previously described isolates^{2,3}, whereas CBL1 (ref. 4) is similar to both LAV-1 BRU and HTLV-IIIb.

As previously reported^{5–7}, monoclonal antibodies raised against CBL1 gag proteins do not distinguish between CBL1, BRU and IIIB, and gag sequence analysis revealed only minor nucleotide variation. Further sequence analysis provided by P. Highfield of Wellcome Laboratories (personal communication), representing in total 2,443 nucleotides in *env*, *tat* and *nef*, show that CBL1 has 98.0% amino acids in common with LAV-1 BRU and 97.8% with HTLV-IIIb (BH10 clone), whereas the similarity in the same regions between BH10 and BRU is 98.3%. The *tat* sequence was most variable, with 94.2% of the CBL1 sequence identical to both BRU and BH10, 83.7% to SF-2 and 74.4% to ELI.

On account of the close sequence relationship of CBL1 to LAV-1, HTLV-IIIb and the two University of Nebraska strains^{8–10}, I cannot exclude the possibility of cross-contamination during its isolation and subsequent adaptation to grow in T-cell lines — both LAV-1 and HTLV-IIIb were propagated in our laboratory at that time. But note that minor sequence variation can result in marked differences in the biological properties of the Nebraska strains⁹ and also of CBL1 compared with LAV-1 and HTLV-IIIb (refs 11, 12).

ROBIN A. WEISS

Chester Beatty Laboratories,
Institute of Cancer Research,
Fulham Road,
London SW3 6JB, UK

1. Anderson, C., Coles, P. & Concar, D. *Nature* **347**, 603 (1990).
2. Goudsmit, J. *et al.* *AIDS* **4**, 559–564 (1990).
3. Schulz, T.F. *et al.* *J. Virol.* **64**, 5177–5182 (1990).
4. Weiss, R.A., Clapham, P. R., Cheingsong-Popov, R., Dalgleish, A. G., Carne, C. A., Weller, I. V. D. & Tedder, R. S. *Nature* **316**, 69–72 (1985).
5. Ferns, R.B., Tedder, R.S. & Weiss, R.A. *J. gen. Virol.* **68**, 1543–1551 (1987).
6. Spence, R.P., Jarvill, W.M., Ferns, R.B., Tedder, R.S. & Parker, D. *J. gen. Virol.* **70**, 2843–2851 (1989).
7. Spence, R.P. *et al.* *J. gen. Virol.* **70**, 2853–2863 (1989).
8. Golub, E.I., Li, G. & Volsky, D.J. *J. Virol.* **64**, 3654–3660 (1990).
9. Stevenson, M. *et al.* *J. Virol.* **64**, 3792–3803 (1990).
10. Wain-Hobson, S. & Myers, G. *Nature* **347**, 18 (1990).
11. Spickett, G. *et al.* *AIDS Res. Human Retroviruses* **5**, 355–366 (1989).
12. Wraith, M.N., Schimpf, A., Pennington, T.H., Walker, F. & Sewell, H.F. *Clin. exp. Immun.* (in the press).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references. □

Questions of life

Alan Holland

Embryo Experimentation. Edited by Peter Singer. Cambridge University Press: 1990. Pp. 263. £27.50, \$39.50.

The Vatican, the Law and the Human Embryo. By Michael J. Coughlan. Macmillan/University of Iowa Press: 1990. Pp. 125. Hbk £12.99; pbk \$8.95.

WHEN does fertilization of the female human egg occur? Is it the moment when the male sperm first enters the outer membrane of the egg; or about 24 hours later when male and female chromosomes finally fuse? This was the problem facing the legislature of the Australian state of Victoria in 1986 when Alan Trounson and his team sought permission to test for the success of microinjection of a single sperm into an egg before the chromosomes fused, under an act which prohibited the fertilizing of an egg specifically for research purposes. The exploitation of the discovery that fertilization is a process thus forced an amendment to the existing act, whose effect was to rule that an embryo begins to exist only when fertilization is complete. It is a case that represents perfectly the complex of scientific, ethical and legal issues which are the subject of *Embryo Experimentation*.

The editors, who also contribute much of the text, are based at the Centre for Human Bioethics at Monash University in Victoria. Their up-to-the-minute command of the ground comes as no surprise in light of the fact that this is the state which led the world in its legislation on embryo research and *in vitro* fertilization generally, and whose scientists are in the forefront of these activities. The text is clear, and accessible to the nonspecialist. The discussion, although the work of many hands, is marked by a scrupulous determination to claim no more than the argument warrants. The book as a whole mounts as effective an ethical case for embryo research as has yet been made, not least through its careful dismantling of the grounds for objection.

Some of the contributions have already appeared in the medical, legal and philosophical journals, but it is useful to have the interlocking issues clearly displayed in one volume. We learn about the purposes of embryo research, where it is now, and where it is headed; about the variety of arguments over whether fertilization, segmentation or some later stage of an embryo's development constitutes the stage at which it acquires moral status; and

about the disagreements over whether legislation is appropriate and, if so, what form it should take. Among the many interesting points to emerge, whose ramifications have yet to be fully explored, is the problem of how far considerations based upon the embryo *in vivo* can be applied to the case of the embryo *in vitro* — a problem which leads Karen

of being supported independently of religious belief. It thereby abandons the appeal to universal moral principles, which has been characteristic of the Catholic, as distinct from the Protestant tradition. Moreover, it cannot be squared with Thomistic psychology, which holds that rational human nature can only be ensouled in a fairly complex kind of body. If it is to merit the attention of legislators, Coughlan argues, Catholic teaching must return to 'natural reason', which may well mean modifying its stance on embryo legislation and most probably on early abortion too.

Can the doubts about embryo research be regarded as settled, except in so far as they derive from a religious perspective? Not entirely. An unsatisfactory feature of both these books, to my mind, is the

weight that is placed on the argument that the early embryo is not a human being because it is not any kind of individual at all until all possibility of twinning or recombining is past. This argument is also used to deny the embryo certain sorts of potential — by Stephen Buckle to deny its potential to become a human subject, and by Coughlan to deny the ascription to the embryo of a rational nature in the sense in which one might ascribe a flowering nature to a tulip bulb. It is difficult to view this argument as much more than a convenient weapon with which to discomfort the opponents of embryo research — enlivened as it may be by the spectres which twinning allegedly creates of "a single-

parent family within the womb" (Coughlan) or "death with no corpse" (Peter Singer and Helga Kuhse). If twinning and the like simply did not occur, would the case for embryo research really look one whit different? On safer ground is the argument that even if the early embryo is a human being it does not yet have the qualities which bestow moral status; or the point pressed by R. M. Hare, that the banning of experiments would bring no advantage: it would not preserve lives that would otherwise be lost since, without the research, they would not have existed to be preserved.

Many of our long-cherished notions are no doubt due to be upset by recent advances in medical research. But are we quite ready for the shock of 'discovering' that what first appears in the female human's womb upon conception is not a human being? □

Alan Holland is in the Department of Philosophy, Bowland College, Lancaster LA1 4YT, UK.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

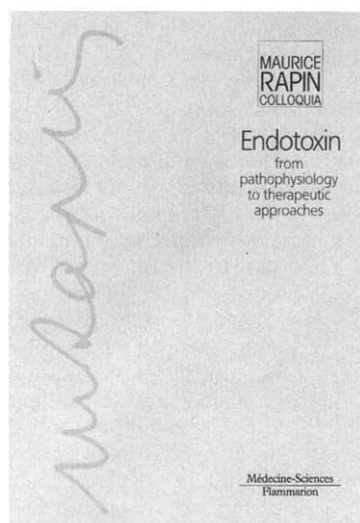
A six-week old embryo — when does it acquire a moral status?

Dawson to the radical doubt as to whether the *in vitro* embryo is indeed an embryo. Another is the undeniable but frequently overlooked fact (to which several contributors draw attention) that women as much as embryos constitute the 'subjects' of embryo research. On the question of regulation, Pascal Kasimba argues persuasively for special legislation, finding no adequate basis for regulation elsewhere, either in the precedents of Anglo-common law, or in the various forms of self-regulation which have been tried or mooted.

Michael Coughlan's monograph *The Vatican, The Law and The Human Embryo*, fills in some of the ethical background, especially concerning the Catholic tradition from which much of the opposition to embryo research derives. Coughlan, too, finds reason to support embryo research, but of particular interest is his contention that the Vatican Congregation's *Instruction* of 1987, declaring that the embryo is a person from the moment of conception, is not capable

A French scientific publisher launches
A NEW SERIES
OF ENGLISH
LANGUAGE BOOKS

Maurice Rapin Colloquia



Dirigé par J. D. Baumgartner,
T. Calandra et J. Carlet
1991. 124 pages. 90 FF

Maurice Rapin Colloquia annually unites a small group of scientists to summarize the latest knowledge in one specific field with, however, a special interest in infection.

This book contains the papers presented during the first meeting dedicated to **Endotoxins and anti-endotoxins**.

This new series of English language books will bring to the reader a unique point of view on some new and advanced aspects of scientific researches.

ORDER FORM

Available from
Librairie Flammarion-4,
Dpt. V.P.C. - 26, rue Jacob
75006 PARIS - FRANCE

NAME _____

ADDRESS _____

Please, send me _____ copy(ies) of:

ENDOTOXINS

Price : 90 FF + 30 FF = **120 FF**

Check enclosed

The return of entropy

P.T. Landsberg

Maxwell's Demon: Entropy, Information, Computing. Edited by H.S. Leff and A.F. Rex. Hilger/Princeton University Press: 1990. Hbk \$75; pbk £17.50, \$24.95.

Complexity, Entropy and the Physics of Information Edited by W.H. Zurek. Addison-Wesley: 1990. Pp. 530. Hbk £40.95, \$48.50; pbk £24.95, \$29.25.

PAST glory and contemporary rejuvenation is the story of entropy. The terms 'aura of unplumbed depth', 'untarnished lustre of novelty' which I applied to it in my 1960 inaugural lecture seem to be valid still. It would, after all, have been excusable had one imagined in 1920 that only pedestrian applications of the entropy concept were to be expected in the future. But one would have been quite wrong.

Past glory is covered in *Maxwell's Demon*. We start with the demon who separates fast from slow molecules, so establishing a temperature difference in what was an equilibrium system, thus violating the entropy law. But the demon has to see the molecules, so photons are needed and energy dissipation through the measuring process exorcises the demon. If instead a one-way trap door is used, dissipation connected with the measurement saves thermodynamics (L. Szilard, 1929; corrected by E. Lubkin, 1987). A new chapter in this history was the entry of entropy into statistics through information theory (C. E. Shannon, 1949). Entropy is now regarded in some contexts as a purely statistical measure (like standard deviation, say) which can be associated with any probability distribution. But the discovery that the demon could be interpreted as a computing automaton reveals some further problems (R. Laing). Both computer and demon must have their memory wiped clean to be in the pristine condition required for the next series of experiments and to achieve a true cycle of operations. This requires the generation of entropy for every bit of information erased (R. Landauer, 1961; C. H. Bennett, 1973), thus making unnecessary the earlier exorcisms. Indeed the whole demoniacal procedure can be approximated as thermodynamically reversible were this erasure to be omitted. All this and more is well told in *Maxwell's Demon*, the editors of which have given us a reliable and well-organized book containing 25 published papers (only four are post-1985). These include all papers mentioned above, except for Shannon's work. I would have preferred a shorter book omitting one or two papers (editors really must be more

severe), but I whole-heartedly recommend it as it stands, for general reading and as a supplement to physics courses.

The contemporary rejuvenation of entropy is exemplified by *Complexity, Entropy and the Physics of Information*, which is replete with modern ideas and matching jargon (glossary not provided).

So to rejuvenation. The percolation of entropy into statistics since the 1950s and into computing since the 1970s is impressive enough. But in addition, the 1970s and 1980s brought increasing preoccupation with the problem of measurement in quantum mechanics and hence consideration of entropy generation in the measurement process. Entropy also entered into the quantum mechanical uncertainty relation (surprisingly, noted only by B. Schumacher and M. H. Partovi). But perhaps the most exciting element of rejuvenation came from classical general relativity in yielding black-hole thermodynamics, when no thermodynamic elements had been inserted into the foundations of the theory. All these aspects are discussed in Zurek's highly stimulating but rather diffuse multi-author book, the usefulness of which would have been increased by improved grouping of the 32 papers into chapters. The book touches on key problems such as the statistical mechanics of the interior of a black hole, complexity, and whether statistical and algorithmic entropy can be combined to yield a deeper insight into the second law of thermodynamics. Both this book and *Maxwell's Demon* discuss the drop in the entropy of a system due to measurement. But if a measurement recognizes a quantum state to be actually a doublet, then the entropy is increased by the measurement. In such cases entropy and 'order' can both increase together.

Of course there is also some nonsense in these papers, and this is expected (because the importance of a subject can be estimated by the amount of nonsense written about it), but this highly stimulating book can be strongly recommended to all (theoretical) physicists.

I end with a few questions, not actually raised in these books, but related to them. What is the effect of gravity on entropy calculations? We need a reference maximum entropy of the universe. Could it be obtained by regarding the whole mass of the universe as concentrated in a single black hole? Is the universe the result of a vacuum fluctuation? Is thermodynamics violated by a macroscopic process which lowers the entropy by a microscopic amount? I expect that the answers must be left to future discussions, for, even on the Maxwell demon problem, they are still in full swing. □

P. T. Landsberg is in the Faculty of Mathematical Studies, University of Southampton, Highfield, Southampton SO9 5NH, UK.

Linked functions

Gary K. Ackers

Binding and Linkage: Functional Chemistry of Biological Macromolecules. By Jeffries Wyman and Stanley J. Gill. University Science Books, Mill Valley, California 94941, USA/WH Freeman: 1990. Pp.330. \$48, £31.95.

BIOLOGICAL macromolecules such as proteins and nucleic acids 'make their living' by forming highly specific complexes with other macromolecules, and small molecular species (for example, enzyme substrates), or by changing their shapes in response to environmental 'signals'. Regulatory effects arise when two or more of these processes exert mutual influences on each other, with the macromolecule acting as a transducer of free energy. The Wyman theory of 'linked functions' provides a predictive framework for understanding these mutual influences — a kind of 'molecular sociology', as Jeffries Wyman has enjoyed calling it during the past four decades. The power of Wyman's conceptual approach is exemplified by its application to the Bohr effect on haemoglobin, a linkage between the two functions of oxygen binding and proton binding (or release). From an experimentally measured number of protons that are released by the haemoglobin molecule upon binding four oxygens, the exact change in oxygen affinity with changes in pH can be predicted. Analogous relationships can be deduced for linked functions in biological macromolecules of every class, including allosteric enzymes, repressor/operator gene control complexes, and membrane receptors.

What makes linkage theory work is a set of special relations (such as the Gibbs-Duhem equation) between the classic thermodynamic potentials. Wyman's ideas, like those of G. N. Lewis, are a creative and powerful extension of these fundamental relationships and should have widespread applicability to biochemical systems. But the linked-functions approach has been under used by researchers and under-taught to research students in spite of a few well-publicized successes such as the haemoglobin case cited above. This is partly attributable to the inherent subtlety and difficulty of thermodynamic reasoning (often not helped by an uninspired exposure of students to the basic subject), and partly to the fact that Wyman's major writings on the subject were frequently opaque to all but the most avid devotees who would take the trouble to decipher the seemingly cryptic, but ultimately practical equations. Thus although the importance of thermodynamic linkage has been recognized by many, its actual under-



Rummaging around — a town's debris is encountered by a polar bear on its migration route in Canada. Taken from *Into Harmony with the Planet: The Delicate Balance Between Industry and the Environment* by Michael Allaby published by Bloomsbury at £14.99.

standing has been achieved by only a few.

Binding and Linkage is a refreshingly readable monograph that will be useful both to graduate students and more senior researchers. The first three chapters provide a clear introductory background to the more advanced theory covered in later chapters. Because structure-energy correlations lie at the heart of mechanistic understanding, a gallery of macromolecular structures is presented at the outset to illustrate the organizational levels and types of functionally related interactions of interest. Although linkage relationships apply, in principle, at any level of structural detail, the theory has special advantages when the interacting subsystems are larger than atomic scale. Applications are thus focused on processes such as the associations between whole subunits (with their sometimes dramatic influence on the binding of ligands), the energetic coupling between conformational 'melting' of intramolecular domains, and the mutual coupling between ligands which bind at widely separated sites on the same macromolecule. Using linkage relationships, it is feasible to decipher the rules of functional coupling for molecules so large that atomic-level details are inaccessible or mechanistically uninterpretable. Illustrative examples of linkage phenomena are presented throughout the book from work by numerous authors, including highly significant contributions from Stanley Gill's own research, and from collaboration between Wyman and Gill.

A unique feature of *Binding and Linkage* is the retrospective (in chapter 4), of Wyman's contributions to the early ideas of allosteric regulation. Two mechanistic lines of thought emerged in the attempts to understand cooperativity of haemoglobin oxygen binding and related effects in multi-subunit enzymes. The 1935 model of Linus Pauling, and subsequent exten-

sions by D. Koshland in the 1960s, used the concept of ligand-induced conformation changes of individual subunits, which produced nearest-neighbour interactions within a multimeric protein structure. This 'sequential' concept was also advanced by J. Monod and his associates at the Institut Pasteur in their original (1963) allosteric model. The second, sharply contrasting concept had been proposed by Wyman as early as 1948. Wyman assumed a 'pre-existing equilibrium' between alternative conformational forms of the haemoglobin molecule, with preferential binding of ligands to one of them. This 'concerted transition' concept was incorporated into the famous 'MWC model' paper of 1965, which Monod coauthored with Jeffries Wyman and Jean Pierre Changeux. The crystallographic structure work of Max Perutz, along with numerous solution studies on haemoglobin, have demonstrated unequivocally the existence of two major conformations of the tetrameric molecule with switching properties that have validated Wyman's early conjecture.

Binding and Linkage is much more than a retrospective of the conceptual contributions of Jeffries Wyman to the understanding of regulatory mechanisms in haemoglobin. The later chapters discuss major new ideas and developments of the last decade, and point the way to areas of theory that require additional development. This book offers a very practical set of tools for the experimental biochemist, and also a series of deep insights into the fundamental energetic constraints wherein biological systems must do their functional business. □

Gary K. Ackers is in the Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, Missouri 63110, USA.

Thinking about thought

Douglas Hofstadter

The Creative Mind: Myths and Mechanisms. By Margaret A. Boden. *Weidenfeld & Nicolson/Basic Books*: 1990. Pp.303. £25, \$24.95.

I HAVE long admired Margaret Boden's style and mode of synthesis. Her 1977 book *Artificial Intelligence and Natural Man* (Basic Books) worked marvellously both as a lay-level introduction to artificial intelligence (AI) and as a scholarly overview. A main theme there was that AI can be seen as a quest for the true complexity and depth of the source of our humanity — our minds.

In her new book, Boden focuses on creativity: the most exalted, most human property of the mind. She convincingly argues that creativity is an emergent consequence of billions of coordinated micro-actions in a physical substrate (the brain); that creativity therefore can, at least in principle, be approximated on a computer; and yet that this view need not make us fear being reduced to a bunch of empty, meaningless circuitry. On the contrary, Boden argues, the mechanisms revealed ought to inspire awe at the subtlety and intricacy of the stuff of which we are made.

The strongest aspect of *The Creative Mind* is its discussion of human creativity, rich in insights about the role played by self-imposed constraints, the role of randomness, and creativity's connection with perception and memory. The weakest aspect, in my opinion, is the discussion of several highly vaunted computer models of particular mental processes, in which it is argued that each exemplifies a successful mechanization of some crucial element of creativity, even if on a very small scale.

One of Boden's deepest insights is that creative minds not only work under constraints, but in fact need them: "far from being the antithesis of creativity, constraints on thinking are what make it possible." She gives many examples, including grammar in language and tonal harmony in music. In discussing what might distinguish Mozart from ordinary mortals, she says: "These rare individuals . . . can search — and transform — high-level spaces much larger and more complex than those explored by other people. They are in a sense more free than us, for they can generate possibilities that we cannot imagine. Yet they respect constraints more than we do, not less. Where we can do nothing, or at best mentally toss a coin, they are guided by powerful domain-relevant principles on to promising pathways which we cannot even see."

The notions of 'conceptual spaces' and 'levels' of conceptual spaces are crucial to Boden's vision of creativity. A central thesis is that, to the extent one has explicit maps of one's own mind, one acquires new powers of creativity, and that there are unlimited levels of such self-reflection. She describes provocative experiments demonstrating striking differences between very unself-reflective attempts at creativity by very young children, and far more deeply creative acts by more highly self-reflective children just a few years older.

Among many fascinating examples of human and computer creativity discussed is Greek geometer Pappus' proof that an isosceles triangle's base angles are equal. The straightforward proof involves a construction cutting the triangle into congruent halves, but Pappus' proof involves no construction. The trick is to consider the triangle's mirror image as another triangle, then to prove these two triangles congruent. This devilishly ingenious proof, rediscovered two millennia later by a computer program whose author had never seen it before, seems like a perfect validation of the idea of 'creative computers'. But by keenly dissecting the program and revealing the nature of the conceptual spaces it had at its disposal, Boden deflates the accomplishment, pointing out that, in contrast to Pappus, "it could not recognize the interest of the proof it produced".

I only wish she had done the same for many other programs she discusses. Not that I wish to see AI as a whole deflated, but because it is a field in which results are incredibly hard to judge, great care is needed. There is a well-known effect, called the 'Eliza effect' after Joseph Weizenbaum's famous psychotherapeutic program Eliza, whereby people tend to impute far more depth to an AI program's performance than is really there, simply because the program's output consists of symbols — familiar words or images — that, when humans use them, are rife with meaning, and so when computers use them, it is next to impossible to suppress the unconscious aura surrounding them. Not just the lay public is susceptible to this insidious effect; surprisingly, AI cognoscenti are as well.

Thus one has a celebrated team of researchers claiming their program has precisely simulated the thought processes by which Kepler found his laws of planetary motion; then, in view of the discrepancy of the lengths of time it took their program and Kepler to do so, they feel compelled to invent excuses for Kepler's slowness — he had to sleep, eat and so forth. Never mind that the computer was presented with exactly the relevant pieces of numerical data, exactly the proper set of mathematical tools, and nothing else — in total contrast with Kepler him-

self, who worked in a fog where the boundary between what was relevant and what was irrelevant to the problem of orbits was completely uncertain, in an era when the very notion of scientific truth was a glimmer in one's eye, and belief in an intrinsic connection of mathematics with physical law was a great intuitive leap. Yet Boden devotes pages to this group's work with scarcely a cautionary word, leading her readers inevitably to the belief that much of the essence of scientific thought is captured by this model. Not only does this harm the field of artificial intelligence by building up unrealistic expectations; it also makes scientific reasoning seem far simpler than it is.

Elsewhere, Boden extols a program that allegedly 'understands' how Socrates' ability to evoke new ideas in the mind of an apprentice is like a midwife's role in a birth. When inspected up close, however, the program turns out to have no notion of midwifery, birth, 'ideas', 'mind', or Socrates; it merely has two skeletal charts of logical-dependency relations among a dozen or so formulas made up of uninterpreted symbols, and it searches blindly until it discovers a satisfactory alignment of the parts of the two charts. It is up to humans to read meaning into this act of skeletal-chart mapping; it certainly has no intrinsic connection with the understanding necessary for recognizing this — or any — deep analogy. Yet Boden, naively echoing the program's authors, claims that this program really gets at what understanding literary metaphors is all about.

To be fair, she certainly understands what today's researchers are doing and why. But I feel that she too unsceptically propagates researchers' own descriptions of their programs' achievements. Of course, too cynical an attitude would undermine her theme that important progress is being made in computer modelling of creativity. There is a middle ground, however: yes, the programs she touts provide some important new ideas about what minds are, but they do so collectively rather than individually. That is, although each program probably represents far less than its authors claim, nonetheless, when taken together, AI models do add up to a significant new viewpoint about the mind.

This book is undoubtedly not Boden's last word on minds and machines. She is committed to thoughtfully analyzing thought, and is one of the world's best commentators on these matters. Although I like the basic message of *The Creative Mind*, I hope she will go more deeply into these topics in the future. □

Douglas Hofstadter is at the Center for Research on Concepts and Cognition, Indiana University, 510 North Fess Street, Bloomington, Indiana 47408, USA.

Particle physics after a year of LEP

David J. Miller

The Z^0 boson has changed in the last year from being an exotic rarity to being the tool for a wide range of research. CERN's large electron-positron collider, LEP, produces enough Z^0 particles to allow many predictions of particle physics theory and cosmology to be tested.

THE 'standard model'^{1,2} of particle physics has three independent components: electroweak theory, which integrates electromagnetism with the weak nuclear force; quantum chromodynamics (QCD), the theory of the strong nuclear force; and the existence of three families of quarks and leptons. Quarks are fermions (spin- $\frac{1}{2}$ objects) with fractional electric charge, which are confined by QCD forces inside 'hadrons'—particles such as protons, neutrons and mesons, which feel the strong nuclear force. Leptons are the free fermions: the electron, the muon and the tau, all with one unit of charge; and the neutrinos, which are uncharged. The model is a synthesis of what has been learned about fundamental particles and interactions over the past fifty years; it is called a model (rather than a theory) because parts of it remain purely empirical.

Electroweak theory and QCD incorporate rigorous mathematical field theories, but so far there is no experimental evidence for a 'grand unified theory' which might amalgamate them and account for the values of the more than 20 parameters needed to apply them to the three families of quarks and leptons. The first success of the model was the prediction of the existence of the Z^0 and W bosons, the quanta that transmit the weak interactions. When the neutral-current part of the weak force was discovered in neutrino interactions at CERN (the European particle physics laboratory, in Geneva) in 1972³, it became possible to develop Enrico Fermi's conjecture⁴ that nuclear beta decay was closely connected with electromagnetism. The Z^0 carries the weak neutral-current force in the same way as the photon carries the electromagnetic force. The W^+ and W^- carry the weak charged-current force which causes beta decay. The W and Z^0 bosons were first observed directly in proton-antiproton collisions at CERN in 1983⁵.

LEP produced its first Z^0 particles from electron-positron collisions in August 1989⁶, and rapidly overtook the number of Z^0 s produced since spring 1989 by the Stanford Linear Collider (SLC), in California. More than half a million Z^0 decays have now been analysed by the four LEP experiments, OPAL, ALEPH, L3 and DELPHI—each built and run by a separate collaboration of hundreds of physicists from universities and laboratories throughout the world.

What LEP does

When an electron and a positron collide at high energies they may rebound unchanged (elastic scattering, also called Bhabha scattering) or they may annihilate. If they annihilate they form a short-lived intermediate particle, either a photon or a Z^0 boson, whose mass M_1 is equal to the sum of the energies of the two incoming particles. (With energy in GeV, we define our units for mass by setting the velocity of light $c = 1$, so Einstein's $E = mc^2$ becomes $2E_{\text{beam}} = E_1 = M_1$.) In general, M_1 will be the 'wrong mass' to make either a real massless photon or a real Z^0 particle, but if we sweep the (equal) energies of the two beams up from 43 to 47 GeV we sweep the combined energy through the mass of the real Z^0 at 91.1 GeV. Such a sweep is in fact made by combining results from many different runs of LEP at different

energies. The magnet ring is first filled with counter-rotating beams of electrons and positrons at ~ 20 GeV, taken from a chain of other accelerators and accumulation rings. The magnet currents are then ramped up to a chosen value corresponding to an energy close to 45 GeV, while power is fed into the radio-frequency cavities to accelerate the particles. The vacuum inside LEP is so good that the beams can then be kept at constant energy for more than 10 hours, while the experiments collect data.

Each experiment completely surrounds one of the four collision points where the electron and positron bunches meet. Outgoing particles are detected in every direction, except for cones of ~ 40 milliradians around the beam. The detectors are built in layers. The inner layers take multiple samples along the tracks of charged outgoing particles, measuring their momentum by curvature in a magnetic field and identifying some particle types by their rate of ionization or by Cerenkov radiation. The outer layers are 'calorimeters'—massive absorbers in which primary particles interact and deposit their energy in showers of secondary particles. Electrons and high-energy photons (γ -rays) leave all of their energy in the fine-grained first layers of the calorimeters, whereas protons, neutrons and mesons

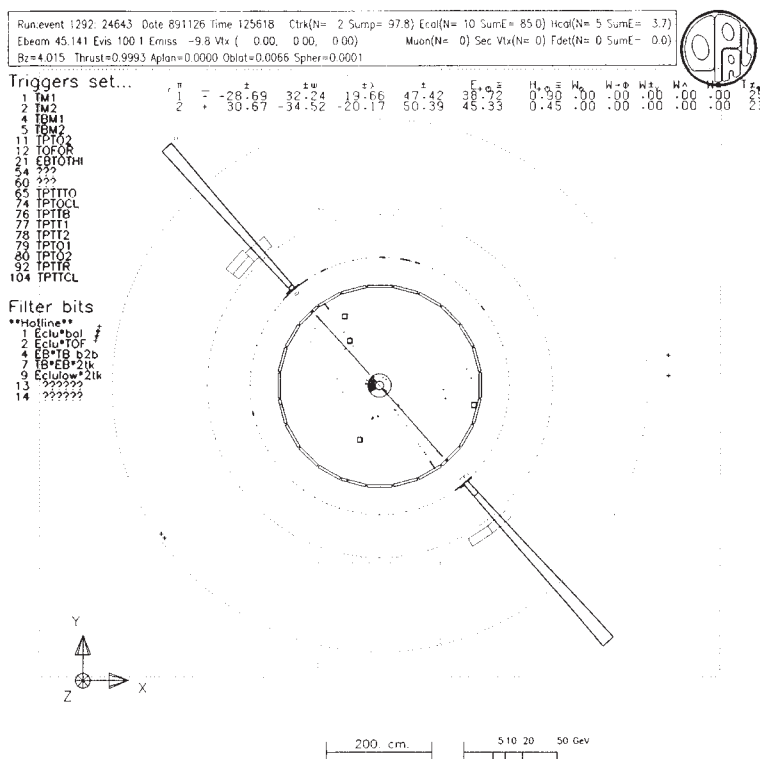


FIG. 1 Computer display of tracks and energy deposition in the OPAL detector for the decay of a Z^0 to e^+e^- . Note the large energies deposited in the electromagnetic calorimeter blocks in line with the tracks (size of box is proportional to energy deposited in the corresponding block). Electrons and positrons are easily recognized because they deposit all of their energy in just a few blocks.

penetrate deeper. High-energy muons penetrate right through all of the calorimeters and are recognized by their tracks in the outer parts of the experiment. Taus decay very characteristically into small numbers of charged particles. Neutrinos do not interact at all.

During the months of running between August 1989 and September 1990 the machine was set at energies from 43 to 47 GeV, with most time spent close to the centre of the Z^0 peak at ~ 45.5 GeV per beam. The cover and Fig. 1 show computer pictures of the tracks and the energy deposited in two typical annihilation events. The characteristic behaviour of each kind of particle allows the identification of four clear categories of annihilation events, which can be counted for each energy. These correspond to the four main visible decay channels of the Z^0 : decays back to electron and positron, to muon and antimuon, to tau and antitau or to a quark-antiquark pair. Their rates at each energy are normalized by comparison with the Bhabha scattering at very small angles (close to the beam direction), which is almost unaffected by Z^0 decays. Figure 2 shows the cross-sections (normalized event rates) for the four channels as a function of the collision mass $M_1 = \sqrt{s} = 2E_{\text{beam}}$. ($s \equiv 4E_{\text{beam}}^2$ is the variable used in calculations because it is a relativistic invariant.) The Z^0 appears as a clear resonant enhancement in all four channels, centred on $\sqrt{s} = M_1 = M_{Z^0}$, with a lineshape that can be fitted using just three parameters: M_{Z^0} , the total width of the Z^0 resonance and the peak cross-section in that channel.

The fitted lineshape parameters and some of the fundamental

quantities that can be derived from them are shown in Table 1. (Definitions are given in the following sections. Readers unfamiliar with the language of Feynman diagrams are strongly encouraged to read Box 1 before proceeding.) The success of the Standard Model is most easily seen by comparing the measurements with the theoretical predictions listed in the last column of the table. Although the results from the four LEP experiments now agree very closely, there were times during 1989/90 when disagreements of two or three standard deviations arose, only to disappear again with the addition of more data. Even where the experimental errors are less than 1% of the number, there is good agreement with theory. In many cases the theoretical uncertainty is larger than the experimental error, owing to (surely temporary) difficulties in computing radiative corrections (see Box 1).

The precision of the data in Table 1 allows us to pin down for the first time such unknown features of the model as the mass of the top quark, as yet unseen, and the number of types of light neutrino—confirming the astronomers' 'standard model' of nucleosynthesis in the Big Bang. Further developments, at LEP and other accelerators, will enable us to measure more of the parameters of the Standard Model (Box 2), and to answer a list of fundamental questions in both particle physics and cosmology.

The Z^0 lineshape

Table 1a lists the parameters of the Z^0 obtained from fits to the energy dependence of the cross-sections for different final states.

BOX 1 HOW TO READ FEYNMAN DIAGRAMS

Quantum electrodynamics (QED)

There is an exact correspondence between individual Feynman diagrams (Fig. B1) and terms in the perturbation series that is used to calculate an interaction amplitude in quantum electrodynamics—the field theory of electricity and magnetism. Each line in a diagram gives a factor in the corresponding term², and the total interaction is given by the sum over all possible diagrams connecting the same external lines. Lines with arrows represent spin- $\frac{1}{2}$ particles (fermions), which carry conserved quantum numbers such as lepton number and quark number. Time runs from left to right, and arrows going backwards in time represent anti-fermions. Wavy lines represent photons—massless bosons with spin 1. They are not conserved and can begin and/or end at interaction vertices between fermion lines. The combination of the incoming and outgoing fermion lines at a vertex is called a 'current', by direct generalization from the electromagnetic current (see Fig. B1 legend), and the photon is the quantized electromagnetic field which interacts with it.

The contribution of a particular term in the perturbation series is proportional to α^n , where α is the fine-structure constant and n is the number of photon-fermion vertices in the corresponding Feynman diagram. Because α is so small ($\alpha \equiv e^2/4\pi\epsilon_0\hbar c \approx 1/137$), higher-order terms (Fig. B1c-e) are treated as small radiative corrections to the simplest tree-level or 'Born' terms (Fig. B1a, b). (But see 'renormalization', below.) Internal lines represent virtual particles, which exist for very short times and are allowed by the Heisenberg uncertainty principle to have different mass values from the corresponding free-particle mass. Such contributions to an interaction amplitude (known as propagators) are multiplied by a factor of approximately $1/(\text{'wrongness of mass' squared})$.

QED amplitudes have a property called 'gauge invariance', which is carried over from the non-uniqueness of the value of the vector potential

in classical electromagnetism. It appears as an arbitrary phase in particle wavefunctions which can be changed without changing any physically measurable quantities. Such a phase change—or 'gauge transformation'—is performed by acting on the wavefunction with an operator which is a generator of a mathematical group called U(1). This is the ►

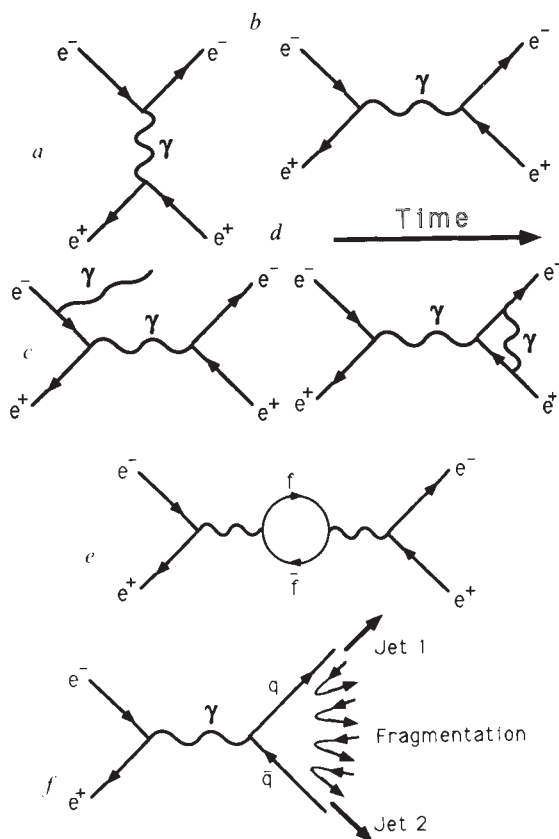


FIG. B1 QED processes. Each diagram of a-e corresponds to a term in the perturbation series for the amplitude for Bhabha scattering (e^-e^- elastic scattering). a and b are the 'tree-level' terms, which make the leading contributions. a is directly connected with the classical expression for the interaction of two currents (e^- to e^- and e^+ to e^+) via the electromagnetic field (quantized as the exchanged photon). c has an extra radiated photon from the incoming electron. d and e both have internal loops which would give infinite contributions but which are cancelled out by renormalization ($f\bar{f}$ may be any charged fermion-antifermion pair, including a pair of f quarks). f is e^+e^- annihilation via a photon to produce a $q\bar{q}$ pair with fragmentation (see Figs B4 and B5). f is the dominant process at lower-energy e^+e^- machines, but is swamped by the Z^0 resonance at LEP.

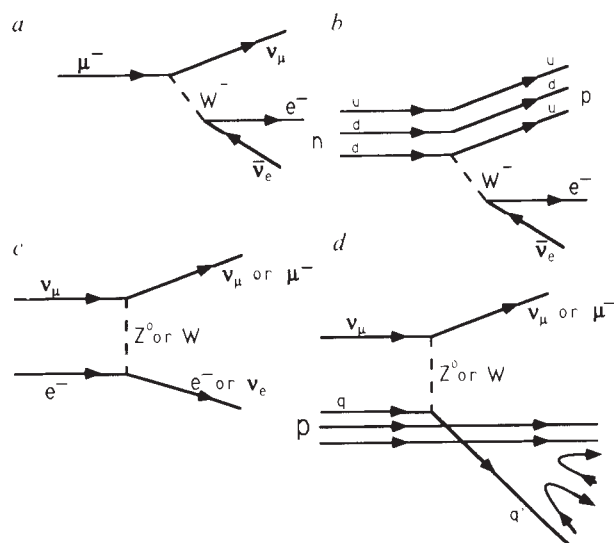


FIG. B2 Weak processes, mediated by the intermediate vector bosons W or Z^0 . *a*, Muon decay, in which the W couples to the two charged currents (μ^- to ν_μ and $\bar{\nu}_e$ to e^-) with the full strength G_μ (see Box 2). *b*, Neutron β -decay, free or inside a nucleus. The $d \rightarrow u$ current is coupled to the W with slightly reduced strength, according to the Cabibbo-Kobayashi-Maskawa theory (Box 2). *c*, Elastic neutrino scattering from an electron. This can be either a charged-current process (with W exchange) or neutral-current (with Z^0 exchange). *d*, Inelastic neutrino scattering from a proton, via charged- or neutral-current interactions, giving a recoil lepton and with fragmentation to give two jets of hadrons (see Fig. B4 legend). Note that one of the jets will contain a proton or neutron—three parallel quark lines.

► 'gauge group' of QED and it is an 'abelian' group—the generators commute with one another, with the physical consequence that three photons cannot interact at one vertex; that is, one photon cannot split into two photons. Both QED and QCD (see below) conserve parity, which means that all scattering or decay angular distributions are indistinguishable from a mirror reflection of themselves. It also means that photons couple equally to the left- and right-handed components of interacting fermions—fermions with their spin backward or forward with respect to the direction of motion.

Electroweak theory, W and Z^0 bosons

Processes governed by the weak interaction, such as muon decay (Fig. B2*a*), neutron β -decay (Fig. B2*b*) and neutrino scattering (Fig. B2*c, d*), can also be represented by Feynman diagrams, using electroweak theory. This is an extension of QED based on the gauge group $SU(2) \times U(1)$, with intermediate massive spin-1 Z^0 and W bosons in addition to the massless photon. The processes are 'weak' at low energies because the damping term due to 'wrongness of mass squared' is of the order of $1/M_W^2$ or $1/M_Z^2$. But when electrons and positrons collide with a centre-of-mass energy close to the Z^0 mass (Fig. B3*a*) the 'wrongness of mass' goes through zero and there is a strong resonance whose height and width are directly linked to the lifetime of the Z^0 (See text; in equation (1) the term $(s - M_Z^2)$ in the denominator is the 'wrongness of mass squared'). Here, the electroweak propagator has produced a formula for the cross-section which is very closely related to classical resonance in forced, damped, simple harmonic motion.)

$SU(2) \times U(1)$ is a non-abelian group—the generators do not commute, which means that the bosons can interact directly with one another. A direct test of the theory, and one of the ultimate goals of LEP, is to prove that the $Z^0 WW$ vertex (Fig. B3*c*) exists, by studying the rate of $W^+ W^-$ production at energies above $2M_W \approx 160$ GeV.

Because the W bosons carry one unit of charge, to conserve charge they have to couple to charge-changing fermion currents, called charged currents. These couplings violate parity conservation maximally—the great discovery of the mid-1950s³⁶. The probability distribution for the directions of the final-state particles in β -decay (from graphs like Fig. B2*b*) can be totally antisymmetric under mirror reflection. The W boson behaves like an equal mixture of vector and axial vector particles, and couples only to the left-handed components of the interacting fermions. (In quantum mechanics the wavefunction of a spin-1 particle like the W or the Z^0 must have the same spatial transformation properties as either a vector, an axial vector or a mixture of the two. In three-dimensional space, a vector (such as momentum) and an axial vector (such as angular

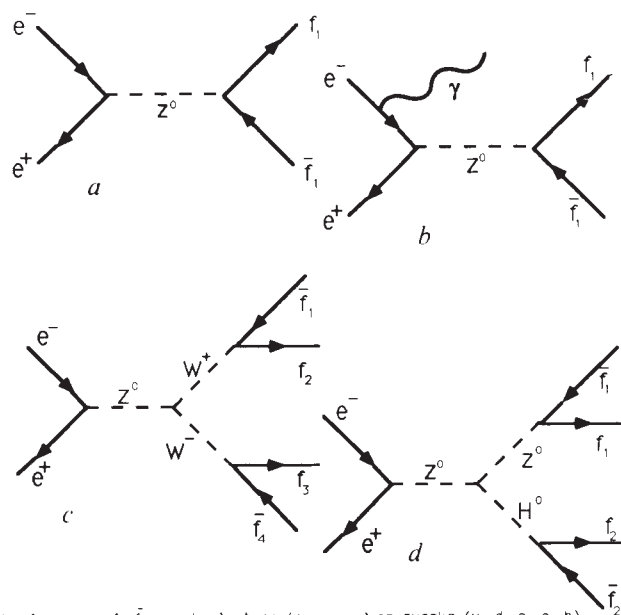
momenta) that are initially parallel have opposite directions after mirror reflection.) The Z^0 , being neutral like the photon, couples to neutral currents of fermions; that is, it mediates interactions that do not change the fermion charge. Unlike the photon, the Z^0 can also couple to a neutral current of neutrinos (see, for example, Fig. B2*c, d*). The Z^0 behaves like an unequal mixture of vector and axial vector, with different vector coupling strengths to different types of fermion, according to their charges. See Table 1 for the measured and predicted values for g_{V_f} and g_{A_f} , the vector and axial couplings to the charged leptons.

Renormalization

Before calculations can be done, field theories have to be 'renormalized' to remove the effects of pathological higher-order terms, which disobey the simple rule that higher-order terms are smaller. Some classes of terms make an infinite contribution to the perturbation series². One set of these terms—called the 'infrared divergence'—corresponds to summing over very-low-energy radiated photons in graphs like Fig. B1*c* or B3*b*. Others involve loop corrections (Fig. B1*e*) and vertex corrections (Fig. B1*d*). Renormalization complicates the theory enormously, but its effects do not destroy the validity of the Feynman diagram language. After the subtraction of infinite correction terms to remove infinite divergent contributions, α becomes an effective coupling which varies as a function of Q^2 , the square of the four-momentum transfer at any photon-fermion vertex. Over the range of Q^2 from a few eV^2 to M_Z^2 , α changes from $\sim 1/137$ to $\sim 1/128$.

Quark diagrams

Some of the quantum numbers carried by the quarks inside ►



The fermions $f, \bar{f}$ may be leptons (e, μ, τ, ν) or quarks (u, d, s, c, b).

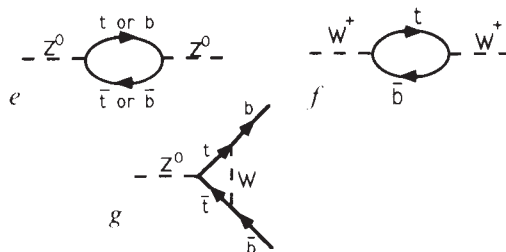


FIG. B3 Z^0 production and decay. *a*, The 'tree-level' term. *b*, Initial-state radiation, which can allow higher-energy $e^+ e^-$ collisions to produce Z^0 particles at close to the peak mass (which is why the curves in Fig. 2 have to be corrected before fitting to the simple lineshape formulae, equations (1) and (2)). *c*, The triple-boson vertex, which will be studied when the LEP energy has been raised to more than $2M_W$. *d*, The most likely Higgs boson (H^0) production mechanism at LEP. *e-g*, Some of the higher-order loops involving the t quark—not cancelled in renormalization because its mass is so great.

► strongly interacting particles (hadrons) are conserved by fundamental interactions; others are not. Figure B2b, for example, shows a d quark being changed to a u quark in the β -decay of a neutron, while the other two quarks remain unchanged. No well authenticated hadrons have yet been observed with quantum numbers that could not be generated by either three 'valence' quarks (for baryons such as the proton, neutron, hyperons, etc.) or a valence quark and valence antiquark (for mesons such as the pion, kaon, etc.). We keep track of hadron quantum numbers by drawing the valence quark lines on the Feynman graph.

Quantum chromodynamics (QCD)

QCD is the field theory of the strong forces which confine quarks inside hadrons. Perturbation theory works for only some QCD processes. The non-abelian gauge group SU(3) generates transformations in the 'colour' degree of freedom. Free hadrons are 'colour singlets'. The colour quantum number is carried by the bound quarks and by the gluons (loopy lines on the diagrams)—spin-1 gauge bosons which generate the forces between the quarks. When QCD is renormalized the effective coupling α_s (analogous to α in QED) becomes very large at low Q^2 but falls off logarithmically at higher Q^2 (see text for measurement of α_s at the Z^0 resonance; also see Box 2). This makes it possible to use individual tree-level Feynman diagrams to calculate the amplitudes for high-energy processes such as $Z^0 \rightarrow q\bar{q}$, $Z \rightarrow q\bar{q}g$ (Fig. B4a, b) or deep inelastic neutrino-proton scattering (Fig. B2d), which involve only a few quarks and hard gluons (in this context, 'hard' means high-energy, say ≥ 1 GeV, and 'soft' means low-energy). Such processes are called perturbative because the higher-order terms in the perturbation series are small enough to be neglected. But the soft gluon exchanges between quarks inside a hadron (Fig. B4c) take place at such low energy and Q^2 that $\alpha_s \gg 1$ and the sum of all the graphs in the perturbation series diverges. Results from deep inelastic scattering experiments (Fig. B2d) allow us to probe the distribution functions for the 'partons' inside a proton—the mixture of valence quarks, gluons and a 'sea' of quark-antiquark pairs, all of which together carry the mass of the proton. To calculate the structure of the bound quarks and gluons inside a hadron (Fig. B4c, and worse) it is necessary to develop totally different, non-perturbative techniques, such as lattice gauge theories. These have not yet been able to explain the mass ratios of the hadrons to within 20%, although progress is being made.

Fragmentation

When a quark is given a large momentum, in the transformation $Z^0 \rightarrow q\bar{q}$, for example (Fig. B4a), or in deep inelastic neutrino scattering (Fig. B2d), we believe that it cannot escape alone, but draws out a string of colour 'lines of force' behind it (Fig. B5), analogous to the bundle of magnetic lines of force that would be produced if one pulled apart the poles of a bar magnet. The energy stored in this bundle of lines grows with its length, until it becomes possible to produce a second quark-antiquark pair as a break in the string. The new antiquark combines with the original quark to produce a shorter string—as if the stretched magnet had snapped to make two smaller magnets—and this continues to stretch until another break occurs. Eventually all the energy from the original process is transformed to the energy of the hadrons produced by combinations of quarks and antiquarks from pair production in the stretched string. No free quarks remain and the 'naked colour' of the quarks is concealed. This process is thought to be a consequence of the non-abelian SU(3) group, which gives rise to a triple-gluon vertex (as in Fig. B4c, lower left). Figures B2d and B4a, b show how the fragmentation process is represented in diagrammatic form for sample quark and hard-gluon final states. Gluons carry two colour quantum numbers, so they are connected to two 'fragmentation strings' of hadrons, as in figure B4b. Figure B4a corresponds to a two-jet multi-hadron event at LEP, Fig. B4b to three jets. (Strictly speaking one could fill up the spaces between the quarks and the fragmentation pairs in Figs. B2d and B4a, b with many interconnected soft-gluon lines and quark-antiquark pairs, representing the field that is stretched out (in a colour string) as fragmentation takes place.)

As the snapping of the colour strings in the fragmentation process is essentially random, it is appropriate to try to reproduce it with random-number generators in a computer. Monte Carlo programs³⁷⁻³⁹ incorporating some theoretical constraints from QCD and some phenomenological parameters have been tuned to reproduce the particle multiplicities, momentum distributions and other statistical measures of the final-state hadrons produced at LEP and at lower-energy machines (see, for example, ref. 40). The fact that these programs reproduce the data so well is a non-trivial success for QCD theory. □

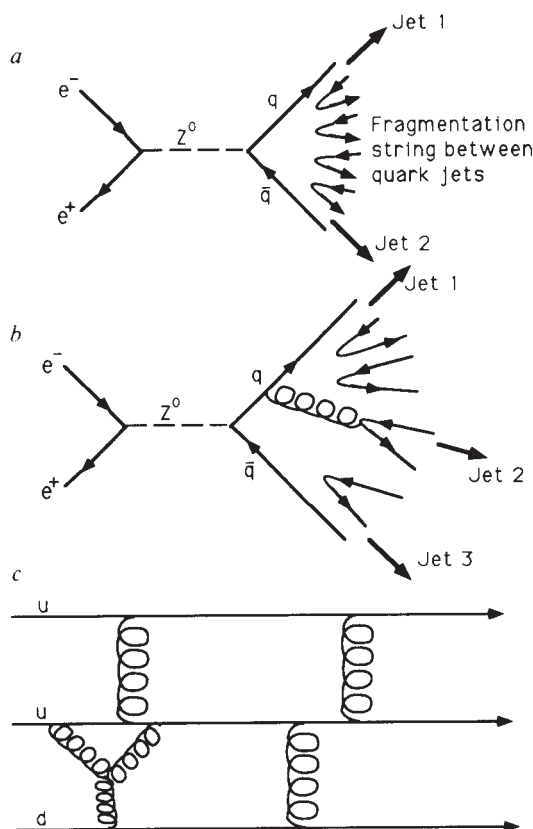


FIG. B4 Fragmentation and jet production. a. The two outgoing quarks are connected by QCD colour lines of force, which snap to produce quark-antiquark ($q\bar{q}$) pairs (as shown in Fig. B5). These eventually come together to form colourless mesons. The triangular space to the right of the $Z^0 q\bar{q}$ vertex could be imagined as filled with interacting gluons (g) from which the extra $q\bar{q}$ pairs are produced. b. Radiation of an energetic ('hard') gluon (loopy line), giving a third jet. (Gluons carry two colour quantum numbers, so they connect two fragmentation strings.) The final particles that have come directly from the parent quark or gluon have the highest momenta in the jet. c. An episode in the binding of three valence quarks by gluon exchange in a proton. There is no leading-order 'tree-level' diagram for such a complicated non-perturbative process. Note the triple-gluon vertex at the bottom left, where a gluon coupled to the d quark splits into two gluons coupled to the u quark, demonstrating the non-abelian character of the QCD gauge-group.

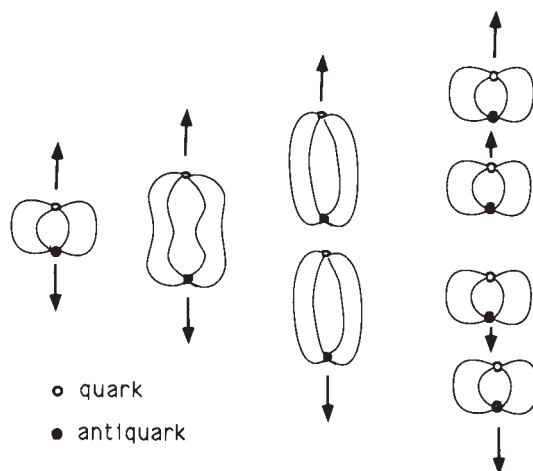


FIG. B5 Fragmentation of the QCD colour field between outward-going quarks and antiquarks (see Box 1 text). As the lines of force are stretched between the primary quarks, which carry colour quantum numbers, the stored energy in the field becomes large enough to produce extra quark-antiquark pairs. These can pair up with oppositely coloured partners to produce 'colourless' hadrons such as pi-mesons, kaons and the like.

TABLE 1 LEP results at the Z^0 resonance and Standard Model predictions

Parameter	ALEPH	DELPHI	L3	OPAL	Average*	Standard Model prediction
(a) Parameters derived from model-independent fits to excitation curves						
M_Z , Z^0 mass (GeV)	91.186 ± 0.013	91.1881 ± 0.013	91.161 ± 0.013	91.174 ± 0.011	$91.177 \pm 0.031^\dagger$	Free parameter (see Box 2)
Γ_Z , Z^0 width (GeV)	2.506 ± 0.026	2.476 ± 0.026	2.492 ± 0.025	2.505 ± 0.020	2.496 ± 0.016	2.500 ± 0.042
σ_0 , Peak hadronic cross-section (nb)	41.78 ± 0.55	42.38 ± 0.96	41.38 ± 0.65	41.88 ± 0.62	$41.78 \pm 0.52^\ddagger$	41.30 ± 0.10
Γ_h , Hadronic partial width (MeV)	1.764 ± 23	1.756 ± 31	1.748 ± 35	1.778 ± 24	$1.764 \pm 16^\S$	1.747 ± 34
Γ_e , Partial width to e^+e^- (MeV)	84.9 ± 1.1	82.0 ± 1.7	84.3 ± 1.4	82.7 ± 1.0	$83.6 \pm 1.0^\S$	83.8 ± 0.9
Γ_μ , Partial width to $\mu^+\mu^-$ (MeV)	80.7 ± 2.2	87.2 ± 3.4	$\parallel$	85.9 ± 2.0	$84.1 \pm 1.4^\ddagger$	83.8 ± 0.9
Γ_τ , Partial width to $\tau^+\tau^-$ (MeV)	81.8 ± 2.2	86.0 ± 3.9	$\parallel$	83.9 ± 2.3	$83.2 \pm 1.5^\ddagger$	83.8 ± 0.9
Γ_ℓ , Mean partial width to charged lepton pairs (MeV)	84.2 ± 1.0	83.7 ± 1.4	83.2 ± 1.4	83.6 ± 0.9	$83.7 \pm 0.7^\ddagger$	83.8 ± 0.9
$R \equiv \Gamma_h/\Gamma_\ell$	20.95 ± 0.30	21.00 ± 0.47	21.20 ± 0.62	21.26 ± 0.31	$21.08 \pm 0.20^\S$	20.86 ± 0.20
Γ_{inv} , Invisible width (MeV) $= \Gamma_Z - \Gamma_h - 3\Gamma_\ell$	489 ± 20	469 ± 27	494 ± 30	476 ± 23	482 ± 16	502 ± 5
(b) Coupling parameters						
$\sin^2 \theta_W(M_Z^2)$ (from Γ_ℓ)	0.2288 ± 0.0023	0.2309 ± 0.0047	0.2272 ± 0.0033	0.2315 ± 0.0024	0.2302 ± 0.0021	Free parameter (see Box 2)
α_s , QCD coupling at $Q^2 = M_Z^2$ (from 3-jet rate)	0.127 ± 0.005	0.114 ± 0.005	0.115 ± 0.005	0.123 ± 0.004	$0.120 \pm 0.012^\parallel$	Free parameter (see Box 2)
$g_V(M_Z^2)$, Z-lepton vector coupling	-0.044 ± 0.008	-0.056 ± 0.021	-0.062 ± 0.018	-0.031 ± 0.015	$-0.045 \pm 0.006^\#$	$-0.0396 \pm 0.012^{**}$
$g_A(M_Z^2)$, Z-lepton axial coupling	-0.502 ± 0.003	-0.502 ± 0.003	-0.497 ± 0.005	-0.501 ± 0.003	$-0.501 \pm 0.002^{\dagger\dagger}$	$(-0.5^\ddagger)$

LEP results are as presented at the 25th International Conference on High Energy Physics, Singapore, 2–8 August 1990. Averages are from review talk by F. Dydak of CERN.

* Averages are quoted with systematic error (1σ) combined with statistical error.

† LEP machine energy uncertainty contributes common systematic error of ± 0.030 GeV, which far exceeds fit errors.

‡ Common theoretical uncertainty of $\pm 1\%$ on Bhabha normalization cross-section for σ_0 , $\pm 0.5\%$ for partial widths.

§ Common theoretical uncertainty of $\pm 1\%$ on subtraction of photon-exchange contribution (Fig. 3a) to Bhabha channel.

$^\parallel$ L3 gives values for the products $\sqrt{\Gamma_e\Gamma_\mu} = 83.3 \pm 1.6$ MeV and $\sqrt{\Gamma_e\Gamma_\tau} = 84.2 \pm 2.0$ MeV.

$^\parallel$ Common systematic error of ± 0.012 from fragmentation uncertainties. Mark II at SLC gets $\alpha_s = 0.123 \pm 0.010$.

$^\#$ Error asymmetric near zero.

** Predicted from $g_V(M_Z^2) = (-\frac{1}{2} + 2 \sin^2 \theta_W)$.

†† Negative sign inferred from neutrino-electron scattering experiments. Common error of ± 0.0015 from error in Γ_ℓ .

‡ $g_A = -0.5$ neglecting radiative corrections.

Figure 2d shows the cross-section $\sigma(s)$ for events with multiparticle jets coming from quark fragmentation (Box 1) when the Z^0 decays to quark-antiquark, as in the event shown on the cover. Figure 2a–c show the results for simpler events in which the final particles are identified as pairs of leptons: large-angle electron-positron (Fig. 1), muon-antimuon or tau-antitau (see Fig. B3a for the Feynman diagram). Close to the Z^0 peak, the points can be fitted, for any pair $f\bar{f}$ of fermions (quarks or leptons) with just three model-independent numbers, M_Z , Γ_Z and σ_f^f :

$$\sigma_f(s) = \frac{S\Gamma_Z^2}{(s - M_Z^2)^2 + S^2\Gamma_Z^2/M_Z^2} \sigma_f^f \quad (1)$$

where M_Z is the Z^0 mass, Γ_Z is its total width, and the peak cross-section σ_f^f is given by

$$\sigma_f^f = \frac{12\pi\Gamma_e\Gamma_f}{M_Z^2\Gamma_Z^2} \quad (2)$$

thao is, proportional to the product of two partial widths, Γ_e for the initial e^+e^- channel and Γ_f for the final $f\bar{f}$ channel. Basic quantum theory requires that $\Gamma_Z = \hbar/\tau_Z$, where τ_Z is the lifetime of the Z^0 . The branching ratio to any particular $f\bar{f}$ channel is $B_f = \Gamma_f/\Gamma_Z$. Five flavours of quark (u, d, c, s and b; up, down, charm, strange and beauty) are produced at the Z^0 resonance, all leading to similar final states containing jets of strongly interacting particles (hadrons), so the fit to $\sigma_h(s) = \sum_{\text{quarks}} \sigma_f(s)$ in Fig. 1d gives $\Gamma_h = \sum_{\text{quarks}} \Gamma_f$. The Z^0 must decay into something, visible or invisible, so the total width is the sum of all the partial widths: $\Gamma_Z = \Gamma_h + 3\Gamma_\ell + \Gamma_{\text{inv}}$, assuming that visible Z^0 decays go only to quarks or to charged leptons, $3\Gamma_\ell \equiv \Gamma_e + \Gamma_\mu + \Gamma_\tau$. The quantity Γ_{inv} covers all Z^0 decays that cannot be detected—the invisible decays. The observed shape of the resonance curve has a well understood correction due to photon radiation in the initial state (Fig. B3b), which reduces the peak by 26% and displaces its position by +110 MeV from the simple

equation (2). (In Table 1, $\sigma_0 = \sum_{\text{quark}} \sigma_f(s = M_Z^2)$ is the corrected, not the observed, peak cross-section.) The e^+e^- final state (Bhabha scattering) is a little more complicated than the others because the extra elastic-scattering graph (Fig. B1a) interferes with the Z^0 decay graph (Fig. B3a)⁷. Nevertheless, all three leptonic partial widths are consistent with being equal to one another, thus confirming ‘lepton universality’—that is, the weak coupling to leptons has the same strength irrespective of the type of lepton. The fitted lineshape parameters from the four LEP experiments are in excellent agreement (Table 1). Results from the Mark II detector at Stanford also agree, but have larger statistical errors, owing to the lower Z^0 production rate at the SLC.

Number of light neutrino species

The electron and muon neutrinos have been well established as distinct particles since 1962^{8,9}. The evidence for the existence of the tau neutrino is very convincing—every observed tau decay has at least one missing neutral particle—even though no one has produced the final proof of seeing taus produced in tau-neutrino interactions. In the Standard Model only decays to neutrino-antineutrino pairs are expected to contribute to the invisible width Γ_{inv} of the Z^0 . Using the best values for the electroweak couplings (Box 2), the partial width for each light neutrino type is predicted to be $\Gamma_\nu = 167.3 \pm 1.6$ MeV. Comparing with the LEP average value for Γ_{inv} in Table 1, we can therefore calculate that there are $N_\nu = 2.96 \pm 0.11$ kinds of light neutrino. An alternative calculation (favoured by F. Dydak in his review of LEP and SLC results at Singapore; see Table 1 footnote for details of conference), assuming lepton universality and using the combination of parameters that minimizes sensitivity to the top quark mass, gives:

$$N_\nu = \frac{\Gamma_\ell}{\Gamma_\nu} \left(\sqrt{\frac{12\pi R}{\sigma_0 M_Z^2}} - R - 3 \right) = 2.89 \pm 0.10 \quad (3)$$

It seems safe to conclude that there are only three kinds of light

neutrino with universal couplings, and to identify them as the neutral partners of the three charged leptons, electron, muon and tau. The value of N_ν is directly relevant to nucleosynthesis in the Big Bang: the number of light neutrino species determined the rate of cooling at the moment when neutrino interactions became so weak that the neutrinos decoupled from thermal equilibrium with the expanding plasma, and the cooling rate in turn determined how many neutrons were available to form light nuclei. Three light neutrinos are completely consistent with current calculations and the observed primordial ^4He , ^2D , ^3He and ^7Li abundances^{10,11}.

The LEP result leaves little room for other contributions to Γ_{inv} , such as neutrinos with non-standard couplings, heavy neutrinos with masses below ~ 45 GeV, or neutral supersymmetric particles with low masses. Of course, new neutral particles could exist with masses greater than half the mass of the Z^0 , as the Z^0 could not decay into pairs of them and LEP could not observe them. There is no interaction yet between LEP results and the two other astrophysical problems involving neutrinos: the low rate of observed interactions of neutrinos from the Sun (reviewed at the Singapore Conference by K. Nakamura, Univ. Tokyo) and the 'dark matter' problem^{11,12}.

Tests of electroweak theory—top-quark mass

The fitted lineshape parameters in the penultimate column of Table 1 are so precise, and so close to the Standard Model predictions, that we can begin to set limits on the corrections to the model due to higher-order processes (Box 1). Higher-order graphs can contain loops (as in Fig. B3e–g) representing particles that are too heavy to be produced as real particles at LEP energy. Small deviations from the standard predictions would therefore be a warning that the model might break down at higher energies. In fact, all that is needed to make the higher-order corrections fit is 'expected unknown physics'—that is, a heavy top (t) quark and, possibly, a heavy Higgs boson (see Box 1).

There is circumstantial evidence for the existence of the t quark. The b ('beauty') quark behaves in every way like an object with charge $-\frac{1}{3}$, with very similar couplings to the d and s quarks. In particular, the angular distribution for production of $b\bar{b}$ events at the TRISTAN e^+e^- collider in Japan is exactly

what is expected for a charge of $-\frac{1}{3}$ (see discussion of b-quark couplings to the Z^0 , below). Not only is it natural to expect the $-\frac{1}{3}$ -charged b quark to have the t quark as a $+\frac{2}{3}$ -charged partner (just as d has u and s has c), but if the electroweak theory is to be renormalized (Box 1), the b quark is in fact required to have a $+\frac{2}{3}$ -charged partner to cancel the infinite contributions from a set of higher-order graphs known as 'anomalies'.

Assuming that the t quark exists, its mass can be estimated by taking it as a free parameter of the model. The best available values are used for α , G_μ (see Box 2) and the fermion masses, with the fitted values of M_Z , $\sin^2 \theta_W$ and α_s as given in Table 1. The measured quantities are not very sensitive to the assumed mass of the Higgs boson, but some of them vary significantly with the top-quark mass m_t , owing especially to the different contributions made by $t\bar{t}$, $t\bar{b}$ and the like in the charged- and neutral-current radiative corrections. (Compare W and Z^0 in Fig. B3e–g. m_b is only ~ 5 GeV, but m_t is probably greater than the masses of the W and Z^0 , so b-quark and t-quark contributions to the loops have very different 'wrongness of mass' effects (Box 1).) Figure 3 shows how $\sin^2 \theta_W$ (Box 2) changes with m_t , in the 'improved Born approximation' scheme^{13–15}. The allowed bands are derived from three different types of experiment. The horizontal band is the direct LEP measurement from the average lepton partial width Γ_ℓ , with no dependence on m_t :

$$\Gamma_\ell = \frac{\alpha M_Z}{12 \sin^2 \theta_W \cos^2 \theta_W} \left[\left(\frac{1}{2} - 2 \sin^2 \theta_W \right)^2 + \left(\frac{1}{2} \right)^2 \right] \quad (4)$$

The narrow line is extracted from the m_t dependence of M_Z , using the value of M_Z from Table 1 as a constraint. The rising band comes from a combined estimate of M_W/M_Z (see Box 2) from $p\bar{p}$ colliders and neutrino scattering experiments. The arrow marked 'CDF limit' at 89 GeV is the 95% confidence limit from a direct search for top quark decays by the CDF collaboration at the Fermilab Tevatron $p\bar{p}$ collider (reviewed by L. Pondrom, Univ. Wisconsin, at Singapore). A rough estimate of $m_t \approx 200$ GeV can be obtained from the LEP constraints alone, but the best value comes from using all three bands, giving $m_t = 137 \pm 40$ GeV.

QCD tests

Jets are seen very clearly in Z^0 decays to hadrons (see cover). These are high-energy processes, compared with the scale parameter $\Lambda_{\overline{\text{MS}}}$ in QCD (see Box 2), so they can be described perturbatively (Box 1), with contributions from the lowest-order two-quark graph (Fig. B4a), single hard-gluon radiation (Fig. B4b) and higher-order graphs with more hard (high-energy) gluons or quark pairs. After fragmentation these processes give two jets of hadrons, or three jets, or four jets, and so on. There is a renormalization problem here too (Box 1), since extra low-energy gluons could be added to any of the diagrams. To allow for this a 'sliding scale' has been worked out to divide any set of hadronic events into two-jet, three-jet, four-jet, five-jet and higher categories, as a function of a parameter y_{cut} which fixes the mass scale for what can be called a distinct jet. The fraction R_3 of events with three clear jets, at a fixed value of y_{cut} , is directly proportional to the QCD coupling strength α_s with only small corrections for fragmentation (Box 1). Figure 4 shows how R_3 varies with centre-of-mass energy $\sqrt{s}$, using data from several e^+e^- colliders: PETRA (DESY, Hamburg) and PEP (Stanford) at 20–44 GeV, Tristram (Japan) at 56 GeV and LEP at 91 GeV. The predicted variation of R_3 (or α_s) with energy is very clearly demonstrated. (It is assumed that the final states from massive virtual photon decays (Fig. B1f) at the lower-energy machines fragment in the same way as do Z^0 decays at LEP.)

The value of α_s , as measured from R_3 , is given in Table 1. Other techniques, using energy correlations of final-state particles, give similar values¹⁶. The scale parameter $\Lambda_{\overline{\text{MS}}}$, which determines how rapidly α_s varies with energy, is not yet so well determined, having values between 120 and 300 MeV. For events

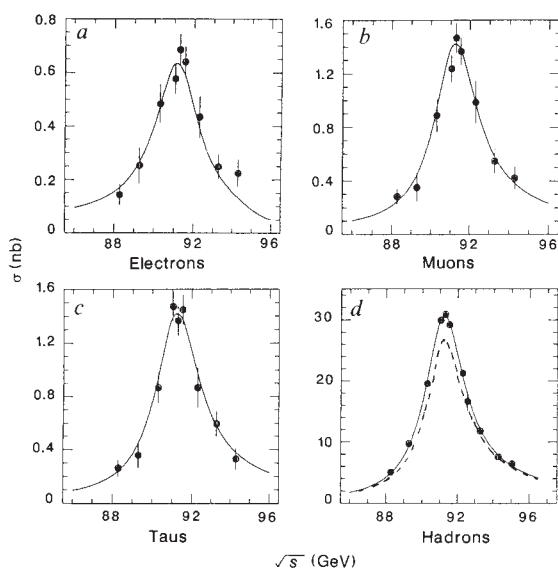


FIG. 2 The Z^0 resonance as seen in e^+e^- cross-sections for the production of pairs of charged leptons or hadrons. (Cross-section σ in nanobarns, where 1 barn = 100 fm².) $\sqrt{s} = 2E_{\text{beam}}$ is the centre-of-mass energy. Solid lines are fits to the Standard Model within three light neutrino types; dotted line in d is the Standard Model prediction with four light neutrino types (OPAL data; other experiments very similar).

BOX 2 THE PARAMETERS OF THE STANDARD MODEL

(For a tabulation of particle properties, together with short reviews of many aspects of particle physics theory including the Standard Model, see ref. 41. For an introductory textbook at postgraduate level see ref. 2.)

Electroweak theory

α , the fine-structure constant of QED. Measured to ten significant figures at low energies.

G_F , the Fermi coupling of the 'charged-current' weak interaction carried by the W boson. Measured in muon decay to six significant figures (see Fig. B2a). The neutral-current coupling is slightly smaller, with strength fixed by the value of $\sin^2 \theta_W$ (see below).

M_Z , the mass of the Z^0 boson which carries the 'neutral current' part of the weak interaction.

M_W , the mass of the W boson. Best value from the CERN and Fermilab pp colliders, $M_W = 79.9 \pm 0.4$ GeV, or via $M_W/M_Z = 0.883 \pm 0.005$ (also from the colliders, reported at Singapore), with M_Z from Table 1.

$\sin^2 \theta_W$. The Weinberg angle θ_W mixes the underlying neutral fields which constitute the photon and the Z^0 (ref. 2). May be estimated experimentally in a number of independent ways, for example from $\sin^2 \theta_W = (1 - M_W^2/M_Z^2)$, from combinations of Z^0 partial widths, or from neutrino neutral-current cross-sections.

M_H , the mass of the Higgs particle⁴², as yet unseen. Experimental lower limit now ~ 45 GeV; see text. If its mass is less than ~ 500 GeV, it provides the simplest explanation for the spontaneous symmetry-breaking effect required to make the electroweak theory renormalizable (see refs. 2, 42).

Parameters of QCD

α_s , coupling constant of QCD. See discussion in text. Value changes with energy, owing to effects of renormalization.

$\Lambda_{\overline{MS}}$, scale parameter (scheme-dependent; see discussion in ref. 41)

which governs the rate of variation of α_s with the square of the four-momentum transfer, Q^2 : $\alpha_s = \text{constant}/\ln(Q^2/\Lambda_{\overline{MS}}^2)$.

Fermion masses

m_e, m_μ, m_τ ; masses of the charged leptons. Well measured⁴¹.

$m_{\nu_e}, m_{\nu_\mu}, m_{\nu_\tau}$; masses of the neutrinos. Consistent with zero, but not required to be zero. (Present limits (Singapore conference and ref. 41): $m_{\nu_e} < 10$ eV; $m_{\nu_\mu} < 0.27$ MeV; $m_{\nu_\tau} < 35$ MeV.)

m_u, m_d, m_s, m_c, m_b ; masses of the known quarks. They cannot be observed as free particles (Box 1), and their effective masses in hadronic matter are not unique (as in the case of an electron in the conduction band of a solid).

m_t , mass of the top quark, as yet unseen. Experimental lower limit of 89 GeV, from the CDF experiment, Fermilab (reported at Singapore). There is strong indirect evidence that it must exist; see text. Top-quark loops (see, for example, Fig. B3e) are the most important source of uncertainties in the radiative corrections to the Z^0 partial widths.

Cabibbo-Kobayashi-Maskawa (CKM) couplings

The mass eigenstates for $-\frac{1}{3}$ -charged d, s and b quarks that are used in building up strongly interacting particles (hadrons) are not the same as the eigenstates that couple via the W to the charged-current weak interaction^{2,40}. The 3×3 CKM matrix connecting the two sets of eigenstates has nine elements, some of which may be complex. It can be parameterized with four real angles, one of which governs the amount of CP violation in K^0 decays³³. The other three determine the rates at which quarks decay into one another (for example, $d \rightarrow u$, Fig. B2b), reducing their effective coupling constant to some fraction of the Fermi constant G_F , defined above. The CKM mixing is applied to the quarks with $-\frac{1}{3}$ charge by convention. Because charged-current processes always connect $+\frac{2}{3}$ to $-\frac{1}{3}$ quarks (Figs. 4d, 5f, g), the mixing could equally well have been applied to the $+\frac{2}{3}$ -charged quarks. □

with four hadron jets it has been impossible to fit the angular distributions without including a very characteristic QCD contribution from the 'triple-gluon vertex' (see Box 1, and Fig. B4c), which splits the single gluon of Fig. B4b into two separate hard gluons¹⁷.

Lepton and quark couplings

Over the past 15 years a wide range of neutrino-scattering and other experiments have attempted to measure g_V , the vector part of the neutral-current coupling to pairs of leptons (see Box 1), but it had never been resolved from zero. Within a year from start-up, the LEP experiments have measured it (Table 1b) and confirmed the electroweak prediction that it should be small and negative. The clearest of the methods, used by all four LEP experiments (reviewed by F. Dydak at Singapore), is to measure the 'forward-backward' asymmetry A_{FB} in the e^+e^- , $\mu^+\mu^-$ and $\tau^+\tau^-$ final states (see Fig. 5). $A_{FB} \equiv (N_+ - N_-)/(N_+ + N_-)$, where N_+ is the number of events in which the positively charged final lepton has a positive component of velocity along the initial positron direction, and N_- is the number with a negative component. At the Z^0 peak, A_{FB} is predicted to be proportional to $(g_V g_A)/(g_V + g_A)^2$. As g_V is very small compared to g_A (see Table 1), A_{FB} at $\sqrt{s} = M_Z$ is expected to be very close to zero. ALEPH has also measured $g_V/g_A = 0.090 \pm 0.030$ by using the forward-backward asymmetry of momentum-weighted hadron charges in jet events.

Three of the experiments (ALEPH, L3 and OPAL) have measured A_{FB} for the production of $b\bar{b}$ pairs at the Z^0 peak. Jets from the heavy b quarks can be 'tagged' by identifying muons or electrons from b decay, which have large transverse momenta with respect to the direction of the jet. (Decays of the lighter u, d, s or c quarks could not produce such high transverse momenta.) The errors are still too large for the results to be worth listing in the table, but the results agree with electroweak theory. A much more definitive measurement has been made of A_{FB} for $b\bar{b}$ production well below the Z^0 peak, at 56 GeV total energy, by the three collaborations (AMY, TOPAZ and VENUS) at the Japanese TRISTAN collider (reviewed by K. Abe, KEK, Japan, at Singapore). At this energy the asymmetry is expected

to be much larger because there is maximal interference between the parity-conserving electromagnetic process (Fig. B1f) and the parity-violating weak process (Fig. B3a), where the Z^0 contribution is greatly reduced due to its 'wrongness of mass'. It is this measured value of $A_{FB} = -0.65 \pm 0.16$ that agrees so well with the Standard Model prediction for a $-\frac{1}{3}$ -charged b quark, and therefore strengthens our belief that the t quark must exist.

As well as measuring asymmetries, the four LEP experiments and Mark II at SLC have measured the product $\Gamma_b B_1$, where B_1 is the branching fraction for b decays to muons or electrons at the Z^0 peak. With a reasonable assumption for the value of B_1 , Γ_b agrees with the electroweak prediction. ALEPH, DELPHI and L3 have recognized a sample of $c\bar{c}$ events and confirmed the predicted value of Γ_c to $\sim 30\%$. OPAL has a measurement,

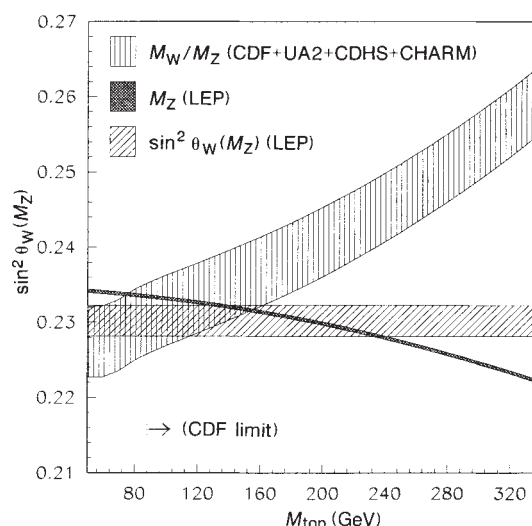


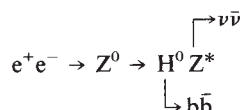
FIG. 3 Constraints on the top-quark mass from measurements of $\sin^2 \theta_W$ (see text), presented by F. Dydak, CERN, at the Singapore conference. The figure assumes a Higgs-particle mass of 200 GeV.

from the observation of hard photons from final-state radiation from quarks, of the total rates for Z^0 to $+\frac{2}{3}$ -charged quarks ($u+c$) and to $-\frac{1}{3}$ -charged quarks ($d+s+b$)—also agreeing with theory to $\sim 25\%$. All of the evidence supports ‘quark universality’: the weak couplings of the $+\frac{2}{3}$ -charged quarks are all equal, as are those of the $-\frac{1}{3}$ -charged quarks, and all are consistent with the electroweak theory.

The search for the Higgs boson

The coupling strength of the Higgs field to any other particle is proportional to the mass of that particle; this follows inescapably from the role of the Higgs field in giving mass to the Z^0 boson in electroweak theory^{2,18,19}. If a Higgs boson exists as the quantum of the Higgs field, and if its mass is less than M_Z , the most probable production process is $Z^0 \rightarrow H^0 + Z^*$ (Fig. B3d), where Z^* is a virtual short-lived Z^0 with the wrong mass, which can decay to e^+e^- , $\mu^+\mu^-$, $\tau^+\tau^-$, quark+antiquark or neutrino+antineutrino. Because its couplings are directly proportional to the masses of the particles it interacts with, the Higgs boson will preferentially decay to the heaviest particles allowed by energy conservation—that is, to photon-photon, if the energy corresponding to the mass of a decaying Higgs is below the e^+e^- energy threshold, to e^+e^- if below the muon-pair threshold, to muons if below the two-pion threshold, to pions if below the kaon-antikaon threshold (that is, too light to produce strange quarks), to kaons if below the charm-anticharm threshold, and to charm if below the beauty-antibeauty threshold.

The LEP experiments have now searched for, and failed to find, Higgs production and decay over the whole mass region from zero up to the machine energy, searching at each energy for the preferred decay modes. As more data accumulate, the search continues to higher Higgs boson masses, above the beauty-antibeauty threshold (at ~ 11 GeV), concentrating on the process



Such events would tend to have two particle-jets, but would differ from normal two-jet Z^0 decays in two ways: the two missing neutrinos would carry off a very large missing momentum, and the presence of b quarks could be recognized by finding electrons or muons with large transverse momentum with respect to the nearest hadron jets (see discussion of ‘tagged’ beauty events, above). The four LEP experiments have all searched for such events, and found none, giving lower limits on the Higgs mass (at 90% confidence level) of 41.6 GeV (ALEPH), 34 GeV (DELPHI), 36.2 GeV (L3) and 39.4 GeV (OPAL). As the LEP programme continues, a Higgs particle should be seen if its

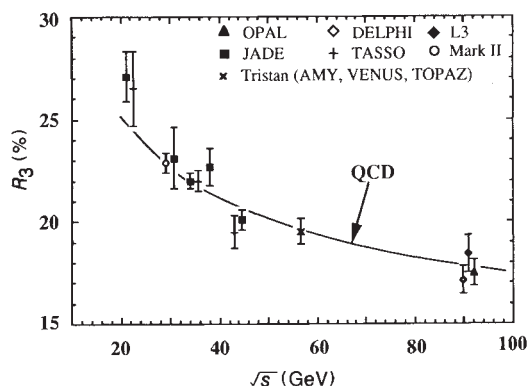


FIG. 4 Variation with energy of the fraction R_3 of three-jet events, for $y_{cut}=0.07$. R_3 is directly proportional to α_s , so these data demonstrate how the coupling ‘constant’ changes with energy, as predicted by QCD (solid line).

mass is ≤ 60 GeV^{18,19}. LEP phase II (planned for 1994), with 100-GeV electrons colliding with 100-GeV positrons, will raise the limit to ~ 80 GeV.

Beyond the Standard Model

The model must be incomplete. There has to be an explanation for the many arbitrary parameters of Box 2. In addition, gravitation has not yet been brought into an experimentally tested relation with the electroweak and QCD theories. Compositeness theories²⁰ attempt to account for the spectrum of quarks, vector bosons and Higgs bosons by introducing ‘preons’, a further underlying set of particles with masses at some ‘compositeness scale’ Λ_c . The LEP and $p\bar{p}$ collider experiments have pushed the lower limit on Λ_c to between 500 GeV and 1 TeV^{21–25}. Supersymmetric theories (SUSY²⁶, superstrings²⁷) have attracted more theoretical support because of the elegant way in which they solve fundamental problems in unifying the component gauge-field theories, and because of the possibility of eventually including gravity. The supersymmetric theories predict a spectrum of integer-spin partners for the observed fermions—‘squarks’ and ‘sleptons’—and half-integer spin partners for the observed bosons—‘Winos’, ‘Zinos’, ‘gluinos’ and ‘photinos’. The LEP lower limits on such particles are all now in the region of 40–45 GeV, at 95% confidence level^{28–31}. ‘Superstring-inspired’ models predict heavier Z' bosons which could modify the neutral-current couplings. But the close fit of the Z^0 lineshape to the Standard Model leaves very little room for extra contributions, and a direct search for heavier Z 's at the Fermilab collider has failed to find any signal up to ~ 190 GeV.

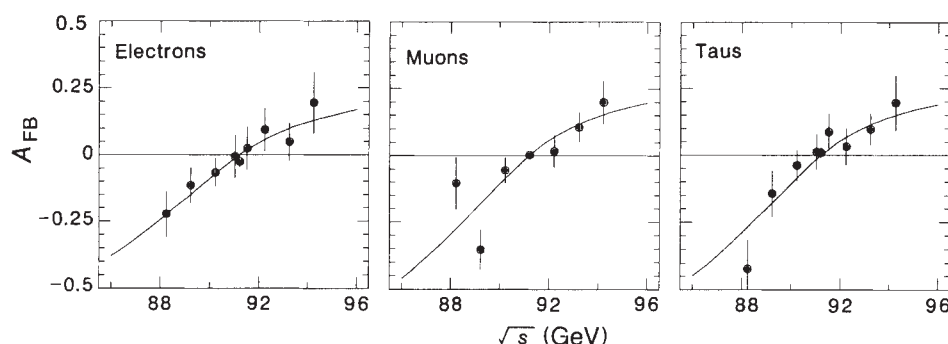
An important question in the physics of the Big Bang is why our Universe has an excess of matter over antimatter¹¹. Sakharov³² suggested a mechanism for this which has subsequently been incorporated into grand unified theories, and which requires violation of CP symmetry (as happens in K^0 decays³³) at a very early stage in the expansion process. The same class of models could also cause protons to decay at a measurable rate, which is why experiments continue in the United States, Japan and Italy to measure the proton lifetime. One model (reported at Singapore by V. Rubakov, Institute for Nuclear Physics, Moscow) requires the Higgs boson mass to be 45 ± 5 GeV—already close to being excluded by LEP.

Future prospects

Experimental techniques at LEP will be refined during 1991 and 1992, and the machine luminosity (essentially, the Z^0 production rate) will increase. There should be considerable improvements in the precision of measuring forward-backward asymmetries, in observing the decays of polarized leptons and in studying the production and decay of heavy quarks. The $b\bar{b}$ channel is especially important, both because it sheds light on the least well measured parameters of the CKM matrix (see Box 2) and because it is particularly sensitive to the mass of the top quark—through $B^0\bar{B}^0$ mixing (a higher-order charged-current effect), through the $b\bar{b}$ production rate and through production asymmetries.

Three distinct sets of improvements to the LEP machine will enable it to test the Standard Model even more comprehensively: ■ Equipment is already installed to measure the transverse polarization of the spins of the electrons and positrons in the beams. This builds up as a result of the interaction of the magnetic moments of the circulating particles with the guiding magnetic fields. If the polarization is significant ($\sim 50\%$), it will be worthwhile to install extra magnets to rotate it into longitudinal polarization at the interaction points, giving ‘left-handed’ or ‘right-handed’ particles whose scattering is very sensitive to the predictions of the model. Once the top-quark mass is known, such measurements with polarized beams might enable us to measure the mass of Higgs boson, even if it is too heavy to be produced at LEP. Stanford is also trying to produce

FIG. 5 Forward-backward asymmetries in lepton pair production. These are generated by the parity-violating interference between vector and axial couplings. Note that the value of A_{FB} passes through zero near the peak of the Z^0 resonance, owing to the small size of the vector coupling g_{V_e} compared with the axial coupling g_{A_e} (see text). To either side of the resonant peak, the vector coupling of the photon interferes with the reduced Z^0 amplitude (Figs. B1b and B3a) to produce larger asymmetries. (In the e^+e^- channel the effects of an extra photon exchange contribution (Fig. B1a) have been subtracted.) OPAL data; other experiments very similar.



polarized beams in the SLC. It should be easier with their linear accelerator and would offset their poorer luminosity.

■ Over the period up to 1995, extra superconducting radio-frequency acceleration cavities will be added to LEP. The goal of this LEP II project is to raise the centre-of-mass energy $\sqrt{s}$ to above twice the mass of the W boson, so that W^+W^- pairs can be produced. The variation of the rate for W pair production with energy will test directly whether electroweak theory contains the expected WWZ triple-gauge-boson coupling (Fig. B3c).

■ Once the extra radio-frequency cavities are in place it will also be possible to increase the number of circulating electron and positron bunches in LEP—thereby further increasing the luminosity and giving much better statistics at the Z^0 resonance. As well as permitting more precise asymmetry measurements to be made, this would turn LEP into a ‘beauty factory’, where millions of beauty-antibeauty pairs will be produced, allowing even better measurements to be made of the badly constrained parameters in the CKM matrix.

Three further clear goals can already be defined which LEP may never achieve:

■ To tie down the last parameter in the CKM matrix, it will be necessary to measure the rate of CP violation in beauty decays. But 10 to 100 times more beauty-antibeauty pairs will be needed than LEP can ever produce. Proposals have been made to build a dedicated beauty factory to do this, with e^+e^- collisions at just over 10 GeV total energy, where the upsilon resonances ($b\bar{b}$ bound states) give a strong enhancement of the production cross-section and decay almost entirely into B mesons.

■ The mass of the top quark is expected to be measured at the Fermilab Tevatron, but only a few events will be seen, complicated by the fragmentation products from spectator quarks left over from the colliding proton and antiproton. Top quark production events will look very different from charm or beauty

production. As the CDF limit $m_t > 89$ GeV requires the top mass to be more than the sum of the W mass and the mass of a beauty particle, a top quark will ‘fall apart’ rapidly into a real W and a b quark (not a ‘weak’ decay because no ‘wrongness of mass’). There will thus be no narrow long-lived T mesons analogous to the K (strange), D (charmed) or B (beauty) mesons. If $m_t > 150$ GeV, neither will there be any ‘toponium’— $t\bar{t}$ bound states analogous to the charmonium $c\bar{c}$ states, J/ψ and ψ' , or the $b\bar{b}$ upsilon states. But if toponium does exist, the exact position of the broad resonant peaks will tell us about the strength of the non-perturbative QCD interquark forces at distances significantly less than 0.1 fm (ref. 34), compared with the region up to 1 fm which has been investigated with charmonium and the upsilons. Accelerator designers are trying³⁵ to develop techniques for building high-luminosity linear e^+e^- colliders to go above the LEP II limit of 100 GeV on 100 GeV, in order to study clean $t\bar{t}$ production, and then to go on towards the 1-TeV region, but there are many technical problems still to be solved.

■ If the mass of the Higgs particle is >80 GeV it must be sought at a higher-energy machine than LEP—which probably means at a hadron collider, because of the unsolved problems in the design of e^+e^- linear colliders. Two very-high-energy machines have been proposed, the Superconducting Supercollider in Texas (pp —not $p\bar{p}$ —at 20 TeV on 20 TeV) and the Large Hadron Collider at CERN (pp at 8 TeV on 8 TeV, with very high luminosity). The search for a heavy Higgs boson is the prime target for both machines. If it does not exist, then something else must happen at an energy scale of 1 TeV to provide an alternative to the Higgs boson in generating the mass of the Z^0 . Some of the possibilities could give very spectacular experimental results, with important implications for cosmology. □

David J. Miller is in the Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK.

- Ellis, J. *Nature* **340**, 277–280 (1989).
- Halzen, F. & Martin, A. D. *Quarks and Leptons* (John Wiley, New York, 1984).
- Hasert, F. J. *et al. Phys. Lett.* **146**, 138–140 (1973).
- Fermi, E. *Nuclear Physics*, 74–84 (University of Chicago Press, 1949).
- Rubbia, C. *Rev. mod. Phys.* **57**, 689–722 (1985).
- Nature* **340**, 677 (1989).
- Greco, M. *et al. Phys. Lett.* **B177**, 97–105 (1986).
- Darby, G. *et al. Phys. Rev. Lett.* **9**, 36–44 (1962).
- Steinberger, J. in *Pions to Quarks, Particle Physics in the 1950s* (eds Brown, L. M., Dresden, M. & Hoddesdon, L.) 307–330 (Cambridge University Press, 1989).
- Steigman, G., Schramm, D. N. & Gunn, J. E. *Phys. Lett.* **B66**, 202–205 (1977).
- Turner, M. S. in *Intersection between Elementary Particle Physics and Cosmology* (eds Piran, T. & Weinberg, S.) 100–214 (World Scientific, Singapore, 1986).
- Smith, P. E. & Lewin, J. D. *Phys. Rep.* **187**, 203–280 (1990).
- Marciano, W. & Sirlin, A. *Phys. Rev.* **D22**, 2695–2717 (1980); *Nucl. Phys.* **B189**, 442–460 (1981).
- Altarelli, G., Kleiss, R. & Verzegnassi, C. (eds) *Z⁰ Physics at LEP Vol. 1*, CERN preprint no. 89/09 (CERN, Geneva, 1989).
- Kennedy, D. C. & Lynn, B. W. *Nucl. Phys.* **B322**, 1–54 (1989).
- Altarelli, G. *Rev. nucl. Part. Sci.* **39**, 357–406 (1989); QCD and Experiment, CERN preprint no. TH 5760/90 (CERN, Geneva, 1990).
- Akrawy, M. *et al.* CERN preprint no. PPE/90–97 (1990).
- Gunion, J. F., Haber, H. E., Kane, G. L. & Dawson, S. (eds) *The Higgs Hunter's Guide* (Addison-Wesley, New York, 1990).
- Franzini, P. J. & Taxil, P. in *Z⁰ Physics at LEP Vol. 2*, CERN preprint no. 89–08 (eds Altarelli, G., Kleiss, R. & Verzegnassi, C.) 58–118 (CERN, Geneva, 1989).
- Eichten, E., Hinchcliffe, I., Lane, K. D. & Quigg, C. *Phys. Rev.* **D34**, 1547–1566 (1986).
- Decamp, D. *et al. Phys. Lett.* **B241**, 623–634 (1990).
- Decamp, D. *et al. Phys. Lett.* **B236**, 501–510 (1990).
- Abreu, P. *et al. Phys. Lett.* **B241**, 449–458 (1990).
- Akrawy, M. Z. *et al. Phys. Lett.* **B242**, 299–308 (1990).
- Akrawy, M. Z. *et al. Phys. Lett.* **B244**, 135–142 (1990).
- Haber, E. & Kane, G. L. *Phys. Rep.* **117**, 75–263 (1985).
- Brink, L. (ed.) *Unification of Fundamental Interactions* (World Scientific, Singapore, 1989).
- Decamp, D. *et al. Phys. Lett.* **B244**, 541–550 (1990).
- Decamp, D. *et al. Phys. Lett.* **B236**, 86–94 (1990).
- Decamp, D. *et al. Phys. Lett.* **B237**, 291–302 (1990).
- Akrawy, M. Z. *et al. Phys. Lett.* **B240**, 250–260 (1990).
- Sakharov, A. *Soviet Phys. JETP Lett.* **5**, 24–27 (1967).
- Miller, D. J. *Nature* **333**, 799–800 (1988).
- Kuhn, J. H. & Zerwas, P. M. *Phys. Rep.* **167**, 321–403 (1988).
- Wilson, E. J. N. *Accelerators for the Twenty-first Century*, CERN preprint no. 90–05 (CERN, Geneva, 1990).
- Franklin, A., Dalitz, R. H., Fitch, V. L. & Telegdi, V. L. in *Pions to Quarks, Particle Physics in the 1950s* (eds Brown, L. M., Dresden, M. & Hoddesdon, L.) 409–484 (Cambridge University Press, 1989).
- Sjöstrand, T. *Comp. Phys. Commun.* **39**, 347–407 (1986).
- Bengtsson, M. & Sjöstrand, T. *Comp. Phys. Commun.* **43**, 367–379 (1987).
- Marchesini, G. & Webber, B. R. *Nucl. Phys.* **B310**, 461–526 (1988).
- Akrawy, M. Z. *et al.* CERN preprint no. EP/90–48 (1990).
- Aguiar-Benitez, M. *et al. Review of Particle Properties*, *Phys. Lett.* **B239**, 1–516 (1990).
- Aitchison, I. *Phys. World* **2**(7), 29–34 (July 1989).

ACKNOWLEDGEMENTS. I apologize to colleagues in ALEPH, DELPHI and L3 for using only data from OPAL on the cover and in Figs 1, 2 and 5. Data from any of the other three experiments would have served equally well. Many results reported here are unpublished, from the Singapore conference referred to in the Table 1 footnote. New LEP results appear as CERN preprints and are mostly published in *Physics Letters B* or in *Zeitschrift für Physik C*. Copies of CERN preprints and reports may be requested by mail from the Librarian, CERN, 1211 Genève 23, Switzerland.

Involvement of p34^{cdc2} in establishing the dependency of S phase on mitosis

Daniel Broek*, Rachel Bartlett, Kevin Crawford & Paul Nurse†

ICRF Cell Cycle Group, Microbiology Unit, Biochemistry Department, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Mutants of *cdc2*⁺ can disrupt the dependency of S phase on completion of the previous mitosis. By changing the state of p34^{cdc2} it is possible to reprogramme a cell from entering mitosis to undergoing S phase. This leads to the proposal that the cell cycle can be considered a p34^{cdc2} cycle, and has implications for the evolution of life cycles.

THE two major events common to all eukaryotic cell cycles are the replication of chromosomes during S phase and their segregation during mitosis. These events usually occur in a strict sequence: S phase precedes mitosis, and mitosis precedes S phase of the succeeding cell cycle. Blocking S phase prevents onset of the subsequent mitosis, which would be lethal if chromosome replication had not been completed, and blocking mitosis prevents initiation of the subsequent S phase, which would otherwise lead to increases in ploidy. The dependencies of S phase and mitosis are examples of checkpoints that ensure orderly progression through the cell cycle². But in certain developmental situations, chromosome replication and segregation can be uncoupled. For example, successive S phases take place without any intervening mitosis in tissues undergoing polytenisation³. Conversely, during meiosis, two rounds of segregation occur without an intervening round of DNA replication. During the early development of certain embryos, the dependencies of S phase upon mitosis⁴, and of mitosis on S phase⁵⁻⁷, are also uncoupled. Nevertheless, in most cells these dependencies are strictly observed and underlie the basic cyclic nature of cell division events.

In the fission yeast *Schizosaccharomyces pombe*, the p34^{cdc2} protein kinase^{8,9} and its positive regulator p80^{cdc25} (refs 10, 11) both play a part in the dependency of mitosis on S phase¹². Protein kinase p34^{cdc2} is required twice during the cell cycle¹³, just before S phase at the control point called 'start' when the cell becomes committed to the cycle¹⁴, and again in late G2, at the onset of mitosis¹⁵. As cells enter mitosis, p34^{cdc2} is dephosphorylated on tyrosine, leading to activation of its protein kinase activity¹⁶. The cell-cycle timing of this activation is determined by a network of gene products, including p80^{cdc25}. When S phase is blocked in wild-type cells, the subsequent mitosis is also blocked. But when S phase is blocked in mutant cells containing either high levels of p80^{cdc25} or an altered form of p34^{cdc2} that does not require p80^{cdc25} for activation, entry into mitosis is no longer blocked¹². The mutant cells fail to detect that DNA replication is incomplete and initiate mitosis regardless. Work using *Xenopus* egg extracts has also suggested that p34^{cdc2} is involved in the dependency of mitosis on S phase¹⁷. In addition, certain drug treatments can induce entry into mitosis before the completion of DNA synthesis¹⁸, and several other genes have been identified that influence this dependency. These include *rcc1*⁺ (baby hamster kidney cells)^{19,20}, *nimA*⁺ and *bimE*⁺ (*Aspergillus nidulans*)^{21,22}, and *RAD9* (*Saccharomyces cerevisiae*)²³,

but it is not known how their gene products interact with p80^{cdc25} or p34^{cdc2}.

Little is known about the dependency of S phase on completion of the previous mitosis. Studies using yeast mutants blocked in mitosis^{24,25}, and vertebrate cell-fusion experiments²⁶, have indicated that some event late in mitosis must be completed before the succeeding S-phase can be initiated. This dependency of S phase on mitosis is uncoupled in embryos of the *Drosophila* mutant *gnu*²⁷, and in heat-treated *Physarum* cultures²⁸. Several models have been proposed to explain how S phase may be dependent on mitosis^{29,30}. Here we describe the isolation of mutants in fission yeast that disrupt the dependency and undergo an extra round of DNA replication without an intervening mitosis. These mutants are altered in the *cdc2*⁺ gene. Destruction of p34^{cdc2} during G2 results in the mutant cells 'forgetting' their position in the cell cycle. On restoration of p34^{cdc2} function, cells carry out an extra S phase instead of undergoing mitosis. Therefore, in addition to its involvement in the dependency of mitosis on S phase¹², changes in p34^{cdc2} can reprogramme the cell to initiate S phase without completing mitosis. Thus p34^{cdc2} has a central role in both the major dependencies regulating the onset of S phase and mitosis during the cell cycle, suggesting that the cell cycle can be considered a p34^{cdc2} cycle³¹.

Diploidizing mutants

A screen was used to isolate fission yeast mutants that undergo a round of DNA replication without carrying out the previous mitosis. We devised a procedure to select for cells displaying the ploidy increases that would be expected from such mutants (Fig. 1a), using a strain in which diploid but not haploid cells could sporulate, and then selecting for sporulating cells. Haploid, homothallic (h⁹⁰) fission yeast cells switch mating-type, generating progeny that are of both h⁺ and h⁻ mating-types³². Cells of opposite mating-type conjugate, resulting in diploid zygotes heterozygous at the mating-type locus (h⁺/h⁻). These can sporulate, forming asci containing four haploid spores. When an h⁹⁰ strain is defective in the *mam2*⁺ gene required for conjugation³³, diploid formation and sporulation are blocked, but sporulation can still take place if cells become diploid by some other means. This situation was exploited by mutagenizing haploid *mam2*⁻ h⁹⁰ cells at 25 °C and then shifting them to 36 °C to allow expression of putative temperature-sensitive phenotypes inducing diploidization (Fig. 1a). A pulse at 36 °C of limited duration (1-3 hours) was used as it was expected that high increases in ploidy would be lethal. After the 36 °C pulse, cells were transferred to sporulation medium³⁴ at 25 °C. Spores were selected by treatment with the enzyme helicase that kills vegetative cells but leaves spores viable³⁵, and colonies formed from these spores were tested for their ability to diploidize after a pulse at 36 °C. Sixteen mutants were isolated that diploidized after exposure to 36 °C. Microscopic examination established that 14 of these underwent defective mitosis leading to abnormal chromosome segregation. At 36 °C they had phenotypes very similar to the previously described *cut* mutants³⁶ that can generate diploid cells. A similar procedure selecting for higher ploidy cells in budding yeast also yielded a mutant, *esp1*-1, defective in chromosome segregation³⁷. The remaining two strains isolated using our screen were not *cut*-like, and were retained as candidates for mutants undergoing DNA replication in the absence of mitosis.

* Present address: Department of Biochemistry, Norris Cancer Research Institute, University of Southern California, School of Medicine, Los Angeles, California 90014, USA.

† To whom correspondence should be addressed.

TABLE 1 Diploid colonies (%) generated in *cdc2^{ts}* mutant alleles

Strain	Treatment			
	25 °C — 36 °C, 4 h	25 °C — 36 °C, -N, 4h	25 °C — 36 °C, -N, 6 h	25 °C — 36 °C, -N — 49 °C, -N
<i>cdc2-M26h⁻</i>	15	17	87	85
<i>cdc2-33h⁻</i>	5	10	29	86
<i>cdc2-56h⁻</i>	<1	5	5	95
<i>cdc2-48h⁻</i>	<1	13	33	86
<i>cdc2-18h⁻</i>	<1	13	25	86
<i>cdc2-L7h⁻</i>	<1	10	23	92

Several *cdc2^{ts}* strains were grown to mid-exponential phase in YE5S at 25 °C, at which time all strains formed less than 1% diploid colonies. The cells were washed three times in nitrogen-free minimal medium and treated in one of the following ways, before being plated on YE5SPB at 25 °C and scored for the formation of diploid colonies as described for Fig. 1. Column 1, cells were resuspended in minimal medium at 36 °C for 4 h; column 2, cells were resuspended in nitrogen-free minimal medium at 36 °C for 4 h; column 3, cells were resuspended in nitrogen-free minimal medium at 36 °C for 6 h; column 4, cells were resuspended in nitrogen-free minimal medium at 36 °C for 4 h and then incubated at 49 °C for 30 min before being plated out at 25 °C. Diploid formation was also measured in wild-type (972h⁻) cells and the following *cdc* mutants that all arrest in G2: *cdc1-7h⁻*, *cdc6-23h⁻*, *cdc25-22h⁻*, *cdc27-P11h⁻*, *cdc28-P8h⁻*. All these strains showed <1% diploid formation under all conditions. Percentage cell survival was calculated by comparing the number of colonies that formed on the plates with the number of cells that were actually plated (counted using a Coulter counter⁵⁰). For all strains under all conditions cell survival was at least 80%.

Mutants in *cdc2⁺*

The above two strains became highly elongated at 36 °C. This is reminiscent of cell division cycle mutants, and so to test whether they were alleles of previously identified *cdc* genes³⁸ they were crossed to various *cdc^{ts}* mutants. Both mutants were found to be alleles of *cdc2⁺* (ref. 39), and were subsequently cloned using the polymerase chain reaction. On sequencing they were each found to be altered in a single amino acid, in one case Pro 137 to Ser 137 and in the other Ala 177 to Thr 177, the former being equivalent to the previously identified *cdc2-M26* allele and the latter to the *cdc2-33* allele⁴⁰.

We tested the ability of these, and of other previously isolated *cdc2^{ts}* mutant alleles, to diploidize after incubation at 36 °C in normal growth medium. We found that *cdc2-M26* exhibited high levels of diploidization (Fig. 1*b*). Cells with the large size characteristic of diploids were confirmed to be diploids by genetic crosses (see legend to Fig. 1*b*), and by showing that the cells contained a diploid DNA content (Figs 2, 3). After four hours

incubation at 36 °C in normal growth medium, *cdc2-33* showed 5% diploidization, but in the other *cdc2^{ts}* mutants no diploidization was detected (Table 1). We wondered whether subjecting the cells to more extreme conditions might lead to more complete denaturation and degradation of the p34^{*cdc2*} protein, resulting in enhanced levels of diploidization. We tested nitrogen starvation, as this should induce protease activity, and a more severe heat treatment that should cause increased denaturation of temperature-sensitive forms of the p34^{*cdc2*} protein. As shown in Table 1, the other *cdc2^{ts}* mutant alleles showed higher levels of diploidization if they were starved of nitrogen at 36 °C, and even higher levels of diploidization were observed if the mutant strains were starved of nitrogen at 36 °C and then heat-treated at 49 °C for 30 minutes, or for 45 minutes (data not shown, but similar to 30-minute treatment). Under all these conditions, cell survival remained high (above 80%). The number of diploids formed was allele-specific, suggesting that the extent of diploidization was determined by the particular mutant form of

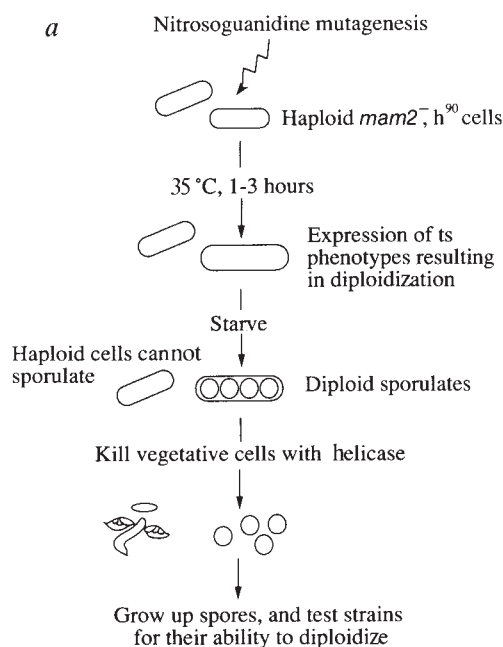
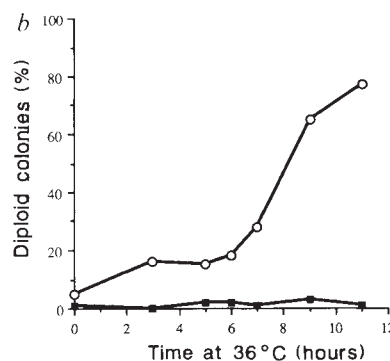


FIG. 1 *a*, Procedure to select for temperature-sensitive fission yeast mutants that increase in ploidy. See text for details. *b*, Diploid colonies generated in the strain *cdc2-M26* (○), and the control strain *cdc25-22* (■), incubated at 36 °C in minimal medium.

METHODS. The strains *cdc2-M26* and *cdc25-22* were grown to mid-exponen-



tial phase in minimal medium (EMM2; ref. 50, as modified by Nurse⁵¹) at 25 °C and then shifted to 36 °C. After various times at 36 °C, cells were plated on yeast extract plus supplements (YE5S: yeast extract 0.5% (Difco); glucose 3%; adenine, histidine, leucine, lysine and uracil all 0.025%) plus 2% agar and 5 mg l⁻¹ phloxin B (ref. 35) (YE5SPB) at 25 °C for two days. Colonies were routinely scored as diploid because of their larger cell size and increased staining with phloxin B on microscopic examination⁵². The cells were confirmed as true diploids by performing a cross between the potential diploid strains *cdc2-M26/cdc2-M26*, *leu1-32/leu1-32*, *h⁺/h⁺* and *cdc2-M26/cdc2-M26*, *leu1⁺/leu1⁺*, *h⁻/h⁻*. Of 22 tetrads, 10 gave a phenotypic ratio (*leu⁺:leu⁻*) of 4:0, 10 gave a ratio of 3:1 and 2 gave a ratio of 2:2. This is consistent with the 5:5:1 ratio of ascus types predicted for a tetraploid cross between true diploids⁵³. Further crosses, involving the centromere-linked marked *lys1-131*, established that within one or two divisions of the diploidization event the centromeres of homologous chromosomes were segregating independently.

TABLE 2 Proportion of *cdc2-33* cells in G1 and G2

Conditions	% G1 cells*	% G2 cells†
25 °C, -N	96	4
25 °C, -C	17	83
36 °C, -N		
25 °C ——— 49 °C, -N	16	84
36 °C, -N ——— 25 °C	89	11

* Do not divide in hydroxyurea.

† Divide into hydroxyurea.

Cells were treated by nitrogen-starvation (row 1), carbon-starvation (row 2), nitrogen-starvation at 36 °C (row 3), and nitrogen-starvation at 36 °C followed by heat-treatment at 49 °C to induce diploidization (row 4). We grew *cdc2-33* cells to mid-exponential phase in minimal medium at 25 °C, washed them in nitrogen-free or carbon-free minimal medium and resuspended them in nitrogen-free or carbon-free minimal medium at 25 °C. After overnight starvation, cells were plated on minimal medium plus 2% agar plates containing 12 mM hydroxyurea (MMHU) and their division patterns followed under the microscope. Most of the cells starved of nitrogen did not divide (row 1) and were thus in G1^{41,42}. Conversely, most of the cells starved of carbon did divide (row 2) and were thus in G2^{41,42}. In a second experiment, *cdc2-33* cells were nitrogen-starved at 36 °C for 4 h as described in Table 1. The culture was then split, with half left at 36 °C (row 3), and half being heat treated at 49 °C for 45 min (row 4). Cells were then plated out on MMHU at 25 °C. Most of the cells plated without prior heat treatment were able to divide (row 3) and therefore in G2; after heat treatment cells were unable to divide (row 4) and therefore in G1. In all cases, control cells plated on minimal medium lacking hydroxyurea underwent between one and two further divisions during the time courses of the experiments.

p34^{cdc2} present in the cells. To check that this phenomenon was specific to *cdc2^{ts}* mutants, other strains were tested for diploidization (Fig. 1, and legend to Table 1). Neither wild-type cells nor other *cdc^{ts}* mutants that block in G2 show any evidence of diploidization. The fact that the only diploidizing mutants of this kind to come out of the screen were *cdc2^{ts}* mutants also supports the idea that this is a specific phenomenon. We conclude that *cdc2^{ts}* mutants can be induced to diploidize by various treatments dependent upon the mutant allele involved, ranging from incubation at 36 °C in normal growth medium to nitrogen starvation at 36 °C combined with heat treatment at 49 °C.

Re-replication of DNA

When *cdc2^{ts}* mutants are incubated at 36 °C, 80–90% of the cells arrest at the late G2 point¹³. This is because a much greater proportion of the cell cycle is located between 'start' and late G2 than between late G2 and 'start'. The diploidization shown by *cdc2^{ts}* mutants can reach 80–90% (Table 1), indicating that the G2-arrested cells undergo diploidization. One possible explanation is that these cells eventually undergo an abnormal mitosis, in which the chromosomes segregate asymmetrically to produce diploid and aploid cells, as is the case for *S. pombe cut* mutants³⁶. To test this possibility, a *cdc2-33* mutant strain was induced to diploidize by nitrogen starvation at 36 °C followed by heat treatment at 49 °C (see Table 1). The cells were plated out at 25 °C, and a microscope was used to monitor the pattern of cell division. Over 95% of the cells produced two viable diploid cells, ruling out the possibility that diploidization was due to asymmetric chromosome segregation, generating diploid and aploid cells.

Several experiments were performed to test the alternative possibility that the *cdc2^{ts}* mutant cells diploidize because they undergo an extra round of DNA replication in the absence of mitosis. First, we analysed whether, when a *cdc2-33* mutant strain was induced to diploidize, the G2-arrested cells re-entered G1 in the absence of mitosis. Whether cells were at a point in the cell cycle before or after DNA replication was determined

by monitoring their ability to undergo cell division in the presence of the DNA synthesis inhibitor hydroxyurea (HU). Cells that have completed S phase and are in G2 can undergo mitosis and divide once in the presence of HU, whereas cells that have not replicated their DNA and are in G1 cannot do so. This was demonstrated using the *cdc2-33* mutant strain grown at 25 °C and starved of nitrogen, leading to G1 arrest, or starved of carbon, leading to G2 arrest^{41,42}. These cells were plated onto growth medium containing HU and the pattern of cell division followed microscopically. As shown in Table 2, the predicted result was obtained with most of the nitrogen-starved G1-arrested cells failing to divide but most of the carbon-starved G2-arrested cells dividing. When the *cdc2-33* mutant strain was incubated at 36 °C and then plated onto growth medium containing HU at 25 °C, 80–90% of the cells were able to divide once (Table 2) indicating that, as expected for a *cdc2^{ts}* mutant, most of the cells were arrested in G2, with a smaller proportion arrested in G1. However, after incubation at 36 °C followed by heat treatment at 49 °C, which induces a high level of diploidization (see Table 1), most of the cells plated onto growth medium containing HU at 25 °C were unable to divide (Table 2), indicat-

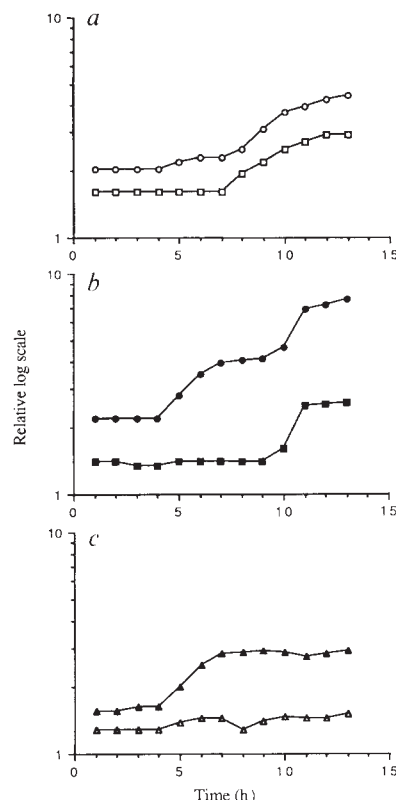


FIG. 2 DNA content (circles), cell number (squares) and DNA per cell (triangles) increases in a *cdc2-33* mutant returned to 25 °C after nitrogen starvation at 36 °C (open symbols), and nitrogen starvation at 36 °C followed by heat treatment at 49 °C to induce diploidization (filled symbols). a, Control cells nitrogen-starved at 36 °C for 4 h; DNA (○), 1 = 25 pg, cell number (□), 1 = 10⁶. b, Cells nitrogen-starved at 36 °C for four hours and then treated at 49 °C for 45 min to induce diploidization; DNA (●), 1 = 20 pg, cell number (■), 1 = 10⁶. c, DNA content per cell for control (△, 1 = 25 fg per cell) and heat-treated (▲, 1 = 20 fg per cell) cells.

METHODS. A *cdc2-33* h⁻ strain was grown to mid-exponential phase in minimal medium at 25 °C and then nitrogen-starved at 36 °C for 4 h as described in Table 1. The culture was split, with half left at 36 °C and half receiving a heat-treatment of 49 °C for 45 min. The cells were then shifted down to 25 °C in minimal medium. DNA content of the cultures was estimated using the diphenylamine method of Bostock⁴¹ with the modification of Gendimenico⁵⁴. Cell number was measured using a Coulter counter⁵⁰. The abscissa scale is arbitrary, with values of cell number and DNA per ml given above.

ing that they were at a point in the cell cycle before DNA replication. This result strongly argues that the mechanism of diploidization involves the *cdc2^{ts}* G2-arrested cells re-entering G1 in the absence of mitosis and cell division.

Next, we wanted to establish whether these cells actually underwent DNA replication in the absence of mitosis and cell division. This was tested by subjecting a *cdc2-33* mutant strain to conditions that induce a high level of diploidization, and then monitoring the DNA content per cell once the cells were returned to growth medium at 25 °C. A culture of *cdc2-33* was nitrogen-starved at 36 °C, so that most of the cells arrested in late G2 (Table 2). The culture was then split in two, with one half receiving a heat treatment at 49 °C to induce diploidization, and the other half remaining at 36 °C as a non-diploidizing control. The cultures were returned to 25 °C at time zero in Fig 2, and DNA content and cell number were monitored. In the non-diploidizing control culture kept at 36 °C, DNA content doubled between 7–11 hours (Fig. 2a), shortly after the cell number increase. This is due to the G2-arrested cells completing mitosis, dividing, and then undergoing S phase of the next cell cycle. DNA content per cell remains constant (Fig. 2c). The pattern of DNA synthesis was quite different in the cells induced to diploidize by heat-treatment at 49 °C (Fig. 2b). In these cells, a doubling in DNA content was observed between 4–7 hours, preceding by about four hours the doubling in cell number at 9–12 hours, and leading to a doubling in the DNA content per cell (Fig. 2c). Microscopic examination of the cells during and after the heat treatment showed no evidence of mitotic events until after the increase in DNA content had taken place. Thus these cells perform a round of DNA replication before undergoing mitosis and cell division. This round of replication appeared to be complete because the DNA content of the *cdc2-33* cells (Fig. 2c), and of *cdc2-56* cells treated in a similar manner (data not shown), essentially doubled to the level seen in diploid cells. We conclude that diploidization of *cdc2^{ts}* mutant strains is due to the G2-arrested cells entering G1 and undergoing an extra S phase without an intervening mitosis.

Finally, we wanted to determine whether the G2-arrested cells re-entered G1 before or after the 'start' control point before DNA replication¹⁴. Because *cdc2⁺* is required at 'start'¹³, we analysed whether *cdc2^{ts}* mutant cells, after treatment to induce diploidization, were then able to undergo DNA replication if they were then maintained at the restrictive temperature for *cdc2^{ts}* instead of being returned to 25 °C. Cells of the *cdc2-M26* strain were nitrogen-starved at 36 °C and then treated at 49 °C to induce efficient diploidization. The culture was then split, with half being returned to 25 °C, a condition that leads to DNA replication and diploidization of *cdc2^{ts}* cells (Fig. 2b). The other half of the culture was incubated at 36 °C. DNA content per cell was monitored for both cultures using a FACS analyser. Figure 3 shows that in the culture returned to 25 °C (left panel), DNA replication starts at ~3.5 hours, leading to a doubling in the DNA content per cell by ~4.5 hours. In contrast, in the culture incubated at 36 °C, there is little evidence of DNA replication by five hours (right panel), even though the growth rate is 50% faster at this temperature. This experiment shows that heat-treated *cdc2^{ts}* cells cannot undergo DNA replication in the absence of *cdc2⁺* gene function. As the *cdc2⁺* function is known to be required at 'start', this result suggests that the cells must traverse 'start' before they can re-replicate their DNA.

p34^{cdc2} disappearance

These experiments establish that cells containing particular temperature-sensitive mutant forms of p34^{cdc2} leave the G2 phase of the cell cycle, re-enter G1, and undergo an extra round of DNA replication after certain treatments, the most effective of which are nitrogen starvation and heat treatment (Table 1). A possible effect of the heat treatment is denaturation of the p34^{cdc2} protein rendering it susceptible to degradation. Nitrogen-starved cells can be expected to be proteolytically active and therefore

to show increased levels of p34^{cdc2} degradation. To test whether these treatments induce diploidization by bringing about p34^{cdc2} degradation, p34^{cdc2} levels were monitored by western blotting of cell extracts prepared from *cdc2^{ts}* and other *cdc^{ts}* mutant strains that arrest in G2^{10,38,43}. Extracts were prepared from cells growing at 25 °C (Fig. 4a, columns 1), conditions under which diploidization does not occur (Table 1), and from cells that were nitrogen-starved at 36 °C and then heat-treated at 49 °C (Fig. 4a, columns 2) to induce diploidization (Table 1). Extracts prepared from cells of three *cdc2^{ts}* mutants before heat treatment contained p34^{cdc2}, but extracts from cells heat-treated at 49 °C to induce diploidization did not (Fig. 4a, A–C, see legend for an assessment of the sensitivity of the western blot). In contrast, extracts prepared from wild-type, *cdc25^{ts}*, *cdc27^{ts}* and *cdc28^{ts}* cells (Fig. 4a, D–G) all contained p34^{cdc2} both before and after heat treatment at 49 °C. Thus, conditions that lead to G2 cells entering G1 in the absence of mitosis are correlated with a drop in p34^{cdc2} levels.

To test this correlation further, a similar experiment was performed using the *cdc2-56* mutant strain which does not diploidize during nitrogen starvation at 36 °C, but does so to a high level after heat treatment at 49 °C (Table 1). During nitrogen-starvation at 36 °C, p34^{cdc2} levels decreased, but the protein was still clearly present (Fig. 4b, panel N). If the cells were then shifted down to 25 °C into fresh medium, only a low level of diploidization was observed and p34^{cdc2} remained present in cell extracts throughout the experiment (Fig. 4b, panel C). In contrast, if the cells were heat-treated at 49 °C before they were returned to 25 °C, there was a high level of diploidization, and p34^{cdc2} became undetectable, only reappearing towards the end of the experiment (Fig. 4b, panel H). Thus, there is a strong correlation between the disappearance of p34^{cdc2} from extracts of G2-arrested cells and their entry into G1, from which they undergo an extra round of DNA replication.

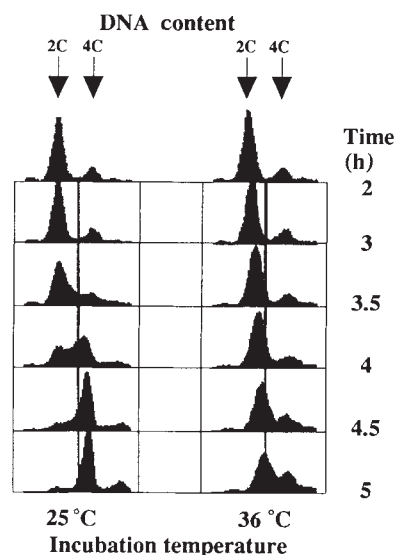
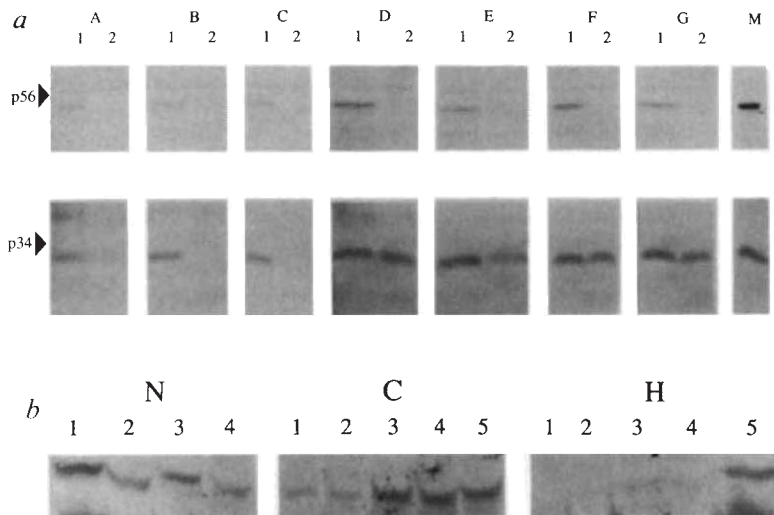


FIG. 3 G2-arrested *cdc2-M26* cells re-replicate their DNA if returned to 25 °C after nitrogen starvation at 36 °C and heat treatment at 49 °C to induce diploidization (left panel), but not if they are kept at 36 °C (right panel). METHODS. *cdc2-M26^h* cells in mid-exponential growth were nitrogen-starved at 36 °C for 4.5 h and shifted to 49 °C for 45 min, essentially as described in Table 1. The cells were transferred to YE5S medium and the culture was split, half being incubated at 25 °C and the other half at 36 °C. Samples were taken at intervals for analysis using a FACSscan (fluorescence-activated cell analyser, Becton-Dickinson). Cells (5×10^7) were fixed in 5 ml 70% ethanol and stored at 4 °C. Immediately before analysis, 1 ml fixed cells were washed in 1 ml 50 mM sodium citrate solution and incubated in 1 ml 50 mM sodium citrate, 100 $\mu\text{g ml}^{-1}$ RNaseA and 5 $\mu\text{g ml}^{-1}$ propidium iodide for 2 h at 37 °C in the dark⁵⁵.

FIG. 4 *a*, p34^{cdc2} and p56^{cdc13} protein levels in various mutant strains in mid-exponential growth at 25 °C (lanes 1), and nitrogen-starved at 36 °C for 4 h and then heat-treated at 49 °C for 30 min (lanes 2). *A*, *cdc2-33 h⁻*; *B*, *cdc2-48 h⁻*; *C*, *cdc2-M26 h⁻*; *D*, 972 h⁻ wild-type; *E*, *cdc25-22 h⁻*; *F*, *cdc27-P11 h⁻*; *G*, *cdc28-P8 h⁻*; *M*, bacterially produced p56^{cdc13} and p34^{cdc2}. *b*, p34^{cdc2} protein levels in *cdc2-56 h⁻* before and after heat treatment at 49 °C. *N*, During nitrogen starvation at 36 °C (1–4; 0, 1, 2, 3 h at 36 °C); *C*, control culture returned to YE5S medium at 25 °C after nitrogen-starvation at 36 °C for 4 h (1–5; 0, 1, 2, 3, 4 h at 25 °C); *H*, culture returned to YE5S medium at 25 °C after nitrogen-starvation at 36 °C for 4 h and heat-treatment at 49 °C for 30 min to induce diploidization. (1–5; 0, 1, 2, 3, 4 h at 25 °C).

METHODS. Strains were grown to mid-exponential phase in YE5S at 25 °C, and then nitrogen-starved at 36 °C for 4 h and heat-treated at 49 °C for 30 min as described in Table 1. Protein samples were prepared essentially as described in ref. 9 and normalized using a BCA assay (Pierce) with the modification of Hill and Straka⁵⁶. SDS-PAGE and electrophoretic transfer of proteins onto Immobilon-P PVDF membrane (Millipore) were performed as described by Laemmli⁵⁷ and Towbin *et al.*⁵⁸ respectively. Membranes were probed with the antibodies PN24, to detect p34^{cdc2} (ref. 8), and SP4, to detect p56^{cdc13} (ref. 9). Proteins were visualized as described by Leary *et al.*⁵⁹, using an alkaline phosphatase-conjugated mouse anti-rabbit monoclonal antibody (Sigma) as the secondary



probe. Dilutions of samples established that the sensitivity of the western blot assay was such that levels of p34^{cdc2} as low as 2% of that normally seen in wild-type cells could have been detected.

Implications

These experiments establish that changing the state of p34^{cdc2} can re-programme a cell from entering mitosis to undergoing S phase. This resetting of cells from G2 to G1 is correlated with a reduction in p34^{cdc2} levels. The *cdc2⁺* gene function is required twice during the cell cycle. The functions can be separated genetically⁴⁴, and the p34^{cdc2} protein differs at its two execution points in terms of its protein kinase activity⁹, its phosphorylation state^{16,45} and its association with other proteins⁴⁶. These observations suggest that p34^{cdc2} can be in at least two states, an S form appropriate for entry into S phase and an M form appropriate for entry into mitosis. We propose that a *cdc2^{ts}* cell arrested in G2 contains M-form p34^{cdc2} that, although inactive, defines the cell to be in the G2 phase of the cell cycle. When p34^{cdc2} is lost from a G2 cell as a consequence of heat treatment, the cell loses the memory that it is in G2, and, on recovery at the permissive temperature for *cdc2^{ts}* mutants, acquires the active S form of p34^{cdc2} leading to entry into S phase rather than mitosis. Our view of the dependencies operating during the cell cycle is illustrated in Fig. 5. Sometime after initiation of S phase, p34^{cdc2} is converted to its M form and, on completion of mitosis, it is converted to its S form. These normal conversions can be short-circuited by mutation to generate a form that can undergo mitosis before S phase is complete¹², or a form that can undergo S phase in the absence of mitosis (this work). Thus, p34^{cdc2} has a central role in both dependencies underlying the basic cyclic nature of cell division events; in fact, the cell cycle can be considered as a p34^{cdc2} cycle³¹.

The levels of p34^{cdc2} remain fairly constant as cells proceed through the mitotic cell cycle⁸, and so the differences between the S and M forms are likely to involve post-translational modifications or associations with other proteins. One candidate is association of p34^{cdc2} with p56^{cdc13}, a cyclin B-like molecule^{47,48} that is present during G2 and is lost as cells

complete mitosis and enter G1. We examined the possibility that association with p56^{cdc13} defines the M form of p34^{cdc2} by monitoring the p56^{cdc13} levels in extracts of *cdc2^{ts}* and other *cdc^{ts}* mutant cells growing at 25 °C, conditions under which cells do not diploidize, and also in cells heat treated at 49 °C to induce diploidization. As shown in Fig. 4*a*, p56^{cdc13} levels fall dramatically in all strains, including those that do not go on to diploidize. Therefore, it is unlikely that p56^{cdc13} association defines the M form, as its association with p34^{cdc2} is lost even in strains that on recovery proceed into mitosis normally. Another candidate for defining the M form of p34^{cdc2} is phosphorylation on Tyr 15 of p34^{cdc2}, which is phosphorylated during G2 but not during G1¹⁶. If this idea is correct, the Phe 15 mutant of p34^{cdc2} that cannot be phosphorylated might be expected to undergo extra rounds of DNA replication. Examination of the mutant revealed no significant tendency to diploidize (data not shown), suggesting that phosphorylation of Tyr 15 is unlikely to fulfil this role. Other modifications and associations are being investigated to establish the pertinent differences between the S and M forms of p34^{cdc2}.

Another implication of this work is that the G1 and G2 periods are interconvertible, being distinguished primarily by the state of p34^{cdc2}. This has also been suggested by others⁴⁹, and if true has implications for the origin of the alternation of haploid and diploid generations during the evolution of life cycles. This is a problem because a primitive single-cell organism would have to evolve simultaneously both conjugation, generating a diploid from fusion of two haploids, and reductional nuclear division, generating haploids from a diploid. One without the other would lead either to increasing ploidy or lethality due to chromosome loss. But if the G1 and G2 periods are interchangeable, a single-cell organism, especially if it had a low number of chromosomes, could evolve mechanisms for interpolating either extra S phases or nuclear divisions into the basic cell cycle, perhaps

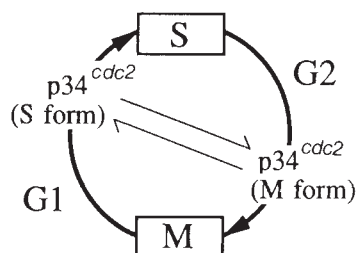


FIG. 5 Dependencies of S phase and mitosis during the cell cycle. Onset of S phase is brought about by p34^{cdc2} (S form) and of mitosis by p34^{cdc2} (M form). Normally, conversion to the M form requires completion of S phase and conversion to the S form requires completion of mitosis. Mutants in *cdc2⁺* can lead to short-circuiting of the normal controls by allowing the generation of a form p34^{cdc2} that can initiate S phase from the G2 phase of the cell cycle, even in the absence of mitosis. Conversely, other mutants in *cdc2⁺* allow a generation of a form of p34^{cdc2} that can initiate mitosis even though S phase is not complete¹².

by manipulating the state of $p34^{cdc2}$. There could be a selective advantage in certain environmental situations for either having low ploidy and consequently being small, or having high ploidy and being large. Once such a system was operative, conjugation

as an alternative mechanism for generating high ploidy, and sporulation as an alternative mechanism for reducing ploidy, could arise independently and thus allow a sexual life cycle to develop. □

Received 6 December; accepted 17 December 1990.

- Mitchison, J. M. in *The Biology of the Cell Cycle*, 201–249 (Cambridge University Press, 1971).
- Hartwell, L. & Weinert, T. *Science* **246**, 629–634 (1989).
- Beermann, W. in *Developmental Studies on Giant Chromosomes*, 1–33 (Springer Verlag, New York, 1972).
- Sluder, G., Miller, F. & Spanjan, K. *J. exp. Zool.* **238**, 325–336 (1986).
- Hara, K., Tydman, P. & Kirschner, M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 462–466 (1980).
- Picard, A., Harricane, M.-C., Labbé, J.-C. & Dorée, M. *Dev. Biol.* **128**, 121–128 (1988).
- Kimelman, D., Kirschner, M. & Scherson, T. *Cell* **48**, 399–407 (1987).
- Simanis, V. & Nurse, P. *Cell* **45**, 261–268 (1986).
- Moreno, S., Hayles, J. & Nurse, P. *Cell* **58**, 361–372 (1989).
- Russell, P. & Nurse, P. *Cell* **45**, 145–153 (1986).
- Moreno, S., Nurse, P. & Russell, P. *Nature* **344**, 549–552 (1990).
- Enoch, T. & Nurse, P. *Cell* **60**, 665–673 (1990).
- Nurse, P. & Bissett, Y. *Nature* **292**, 558–560 (1981).
- Bartlett, R. & Nurse, P. *Bioessays* **10**, 457–463 (1990).
- Nurse, P. *Nature* **344**, 503–508 (1990).
- Gould, K. & Nurse, P. *Nature* **342**, 39–45 (1989).
- Dasso, M. & Newport, J. *Cell* **61**, 811–823 (1990).
- Schlegel, R. & Pardee, A. *Science* **232**, 1264–1266 (1986).
- Nishimoto, T., Eilen, E. & Basilio, C. *Cell* **15**, 475–483 (1978).
- Kai, R., Ohtsubo, M., Sekiguchi, M. & Nishimoto, T. *Molec. Cell. Biol.* **6**, 2027–2032 (1986).
- Osmani, S., Engle, D., Doonan, J. & Morris, N. *Cell* **52**, 241–251 (1988).
- Osmani, S., Pu, R. & Morris, N. *Cell* **53**, 237–244 (1988).
- Weinert, T. & Hartwell, L. *Science* **241**, 317–322 (1988).
- Pringle, J. & Hartwell, L. in *The Molecular Biology of the Yeast Saccharomyces. Life Cycle and Inheritance* (eds Strathern, J., Jones, E. & Broach, J.) 97–142 (Cold Spring Harbor Laboratory, New York, 1981).
- Uemura, T. & Yanagida, M. *EMBO J.* **5**, 1003–1010 (1986).
- Rao, P. & Johnson, R. *Nature* **225**, 159–164 (1970).
- Freeman, M. & Glover, D. *Genes Dev.* **1**, 924–930 (1987).
- Brewer, E. & Rusch, H. *Exp. Cell Res.* **49**, 79–86 (1968).
- Roberts, J. & Weintraub, H. *Cell* **452**, 397–404 (1988).

- Blow, J. & Laskey, R. *Nature* **332**, 546–548 (1988).
- Murray, A. *Nature* **342**, 14–15 (1989).
- Egel, R. in *Molecular Biology of the Fission Yeast*, 31–73. (Academic, New York, 1989).
- Egel, R. *Molec. gen. Genet.* **122**, 339–343 (1973).
- Crandall, M., Egel, R. & MacKay, V. *Adv. Microb. Physiol.* **15**, 307–398 (1977).
- Moreno, S., Klar, A. & Nurse, P. *Meth. Enzym.* **194**, 793–823 (1990).
- Hirano, T., Funahashi, S., Uemura, T. & Yanagida, M. *EMBO J.* **5**, 2973–2979 (1986).
- Baum, P., Yip, C., Goetsch, L. & Byers, B. *Molec. cell. Biol.* **8**, 5386–5397 (1988).
- Nurse, P., Thuriaux, P. & Nasmyth, K. *Molec. gen. Genet.* **146**, 167–178 (1976).
- Nurse, P. & Thuriaux, P. *Genetics* **96**, 627–637 (1980).
- Carr, A., MacNeill, S., Hayles, J. & Nurse, P. *Molec. gen. Genet.* **218**, 41–49 (1989).
- Bostock, C. *Exp. Cell Res.* **60**, 16–26 (1970).
- Costello, G., Rodgers, L. & Beach, D. *Curr. Genet.* **11**, 119–125 (1986).
- Nasmyth, K. & Nurse, P. *Molec. gen. Genet.* **182**, 119–124 (1981).
- Booher, R. & Beach, D. *EMBO J.* **7**, 2321–2327 (1988).
- Moria, A., Draetta, G., Beach, D. & Wang, J. *Cell* **58**, 193–203 (1989).
- Booher, R., Alfa, C., Hyams, J. & Beach, D. *Cell* **58**, 485–497 (1989).
- Hunt, T. *Curr. Op. Cell Biol.* **1**, 268–274 (1989).
- Hagan, I., Hayles, J. & Nurse, P. *J. Cell Sci.* **91**, 587–595 (1988).
- Yoshida, M. et al. *Exp. Cell Res.* **187**, 150–156 (1990).
- Mitchison, J. M. in *Methods in Physiology* Vol. 4, 131–165 (Academic, New York, 1970).
- Nurse, P. *Nature* **256**, 547–551 (1975).
- Kohli, J., Hottinger, H., Munz, P., Strauss, A. & Thuriaux, P. *Genetics* **87**, 471 (1977).
- Roman, H., Phillips, M. & Sands, S. *Genetics* **40**, 546–561 (1955).
- Gendimenico, G., Bouquin, P. & Tramposch, K. *Analyt. Biochem.* **173**, 45–48 (1988).
- Sazer, S. & Sherwood, S. *J. Cell Sci.* **97**, 509–516 (1990).
- Hill, H. & Straka, J. *Analyt. Biochem.* **170**, 203–208 (1988).
- Laemmli, U. *Nature* **227**, 680–685 (1970).
- Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350–4354 (1979).
- Leary, J., Brigati, D. & Ward, D. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4045–4049 (1983).

ACKNOWLEDGEMENTS. We thank Herbert Gutz and Olaf Nielsen for strains, Pauline Jeffreys for technical assistance, and members of the Cell Gutz group for materials, advice and discussions. This work was supported by the ICRF, MRC (P.N.) and NCI (D.B.).

LETTERS TO NATURE

Negative ions in the coma of comet Halley

P. Chaizy*, H. Rème*, J. A. Sauvaud*, C. d'Uston*,
R. P. Lin†, D. E. Larson†, D. L. Mitchell†,
K. A. Anderson†, C. W. Carlson†,
A. Korth‡ & D. A. Mendis§

* Centre d'Etude Spatiale des Rayonnements, 9 Avenue du Colonel, BP 4346, 31029 Toulouse Cedex, France

† Space Sciences Laboratory, University of California, Berkeley, California 94720, USA

‡ Max-Planck-Institut für Aeronomie, Postfach 20, D-3411 Katlenburg-Lindau, Germany

§ Department of Electrical Engineering and Computer Sciences, University of California, San Diego, California 92093, USA

IN March 1986, the Giotto spacecraft encountered comet Halley, approaching to within ~600 km of the nucleus. Results from this encounter have shown that the inner coma contains a mixture of cometary neutral gas and dust, thermal ions and electrons, fast cometary pick-up ions, and decelerated solar-wind ions and electrons, as well as fast neutrals¹ produced by charge exchange between pick-up ions and cold neutrals. Here we report the detection of a new component of the inner coma of comet Halley: negatively charged cometary ions. These ions are observed in three broad mass peaks at 7–19, 22–65 and 85–110 AMU, with densities reaching ≥ 1 , $\sim 5 \times 10^{-2}$ and $\sim 4 \times 10^{-2} \text{ cm}^{-3}$, respectively, at a distance of ~2,300 km from the nucleus. The ion species thought to be present include O^- , OH^- , C^- , CH^- , CN^- and heavier complex CHO molecular ions. As negative ions are easily destroyed by solar radiation at ~1 AU (ref. 2) an efficient production mechanism, so far unidentified, is required to account for the observed densities. The detection of negative ions in the coma near 1 AU implies that in similar neutral gas and dust environments

farther away from the Sun (in Jupiter's or Saturn's magnetospheres³, for example), negative ions should also be present. If the negative-ion densities are large enough, they could play an important part in physical processes such as radiative transfer or charge exchange.

The RPA-1 electron electrostatic analyser⁴ (Fig. 1) on the Giotto spacecraft was designed for measurements of plasma electrons near comet P/Halley (see refs 5–10 for results). Negatively charged ions, however, can also be identified if their thermal and flow speeds are much less than the spacecraft velocity of 68.4 km s^{-1} relative to the comet. Such singly charged cold negative ions will be detected by the RPA-1 EESA in the ram sector with an energy proportional to their mass, M (AMU), $E(\text{eV}) = 5.18 \times 10^{-3} M v^2$, where v is the velocity (km s^{-1}).

Figure 2a shows the spectrum of negatively charged particles at a distance of ~5,800 km from the nucleus. Superimposed is a spectrum from a different sector, representative of all other look directions. At distances of $\leq 20,000 \text{ km}$, the ram sector consistently shows three distinct peaks at energies between ~400 eV and ~3 keV, whereas in all other directions the distribution function decreases monotonically and is near background by ~1 keV. These ram-direction peaks cannot be due to electrons: because the magnetic field from the solar wind is not parallel to the ram direction, these peaks would require three gyrophase-bunched electron beams which maintain the same energy and phase even as the distance from the nucleus varies, and as the direction of the magnetic field changes.

Figure 2b, plotted in units of density assuming the particles are ions, shows that the three ram-sector peaks are completely consistent with cold, negative ions of mass ~7–19, 22–65 and 85–110 AMU (denoted the 17, 30 and 100-AMU peaks, respectively, below) nearly at rest in the comet frame.

Figure 3 shows a spectrogram of the ram-sector count rate as a function of time from ~17,000 km to 2,300 km from the nucleus, where telemetry was lost. Figure 4 shows the radial dependence of the three major peaks. Excluding the spikes,

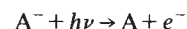
which may be related to dust impacts, the density profiles show $\sim r^{-2.7}$ dependence outside the ionopause and $\sim r^{-4}$ inside. (The ionopause is the boundary of the magnetic-field-free cavity discovered by the Giotto magnetometer¹¹.) No radial dependence is obtained for the 17-AMU peak inside the ionopause because the flux of such ions saturated the detector. The 30- and 100-AMU groups show an increase above the power-law dependence near 10,000 km, approximately where a peak is seen in the positive-ion density ('ion pile-up' region)¹². The negative-ion densities are typically $\leq 10^{-4}$ of the positive-ion densities and $\leq 10^{-6}$ of the neutral densities.

The alternation of high and low count rates every two seconds (Figs 3 and 4), seen most prominently for the ~ 100 -AMU peak but present in all the peaks, is consistent with spin modulation (spin period 4 s). The spin modulation disappears just as the spacecraft crosses the ionopause and enters the magnetic-field-

free cavity. The peak-to-valley ratio for the spin modulation ranges from a few tens of per cent for the 17-AMU peak up to several hundred per cent or more for the 100-AMU peak.

The observed spin modulation can be understood qualitatively as being due to the acceleration of the newly formed negative ions by the motional $(-\mathbf{v}_p \times \mathbf{B})$ electric field (where $\mathbf{v}_p$ is the velocity of the convecting plasma and $\mathbf{B}$ the magnetic field). The lack of spin modulation inside the ionopause then follows naturally from the disappearance of the magnetic field in the cavity.

The primary loss mechanism for most negative ions in the inner Solar System is photodetachment



At 1 AU the lifetime of OH^- against photodetachment is $\tau_{\text{pd}} \approx 0.9$ s (ref. 13). The ion gyroperiod $\tau_g = 2\pi M/qB$ is 19–37 s for singly charged 17-AMU ions in the 30–60 nT magnetic field typical of the inner coma outside the magnetic cavity. Thus, newly formed OH^- ions will just begin their motion under the motional electric field before they are converted to neutrals. The typical velocity of the negative ions will then be $\mathbf{v}_- \approx \mathbf{a}\tau_{\text{pd}}$, where $\mathbf{a} = q(-\mathbf{v}_p \times \mathbf{B})/M$. The observed plasma flow speeds^{14,15} and magnetic fields¹⁶ give negative-ion velocities with a component perpendicular to the ram direction of typically a few tenths of kilometres per second. Such velocities, together with ion temperatures ≤ 350 K (0.030 eV), will produce spin modulation in the detector (see Fig. 1 legend) with peak-to-valley ratios consistent with those observed for the 17-AMU peak. The stronger spin modulation of the 30- and 100-AMU peaks may be due to longer photodetachment lifetimes for the heavier negative ions^{17,18}, together with possibly lower thermal speeds.

Negative ions might be produced from interactions of cometary neutrals with the detector or with the spacecraft by, for example, electron attachment in colliding with the surfaces. From $\sim 2,500$ km to $\sim 40,000$ km the neutral-water vapour

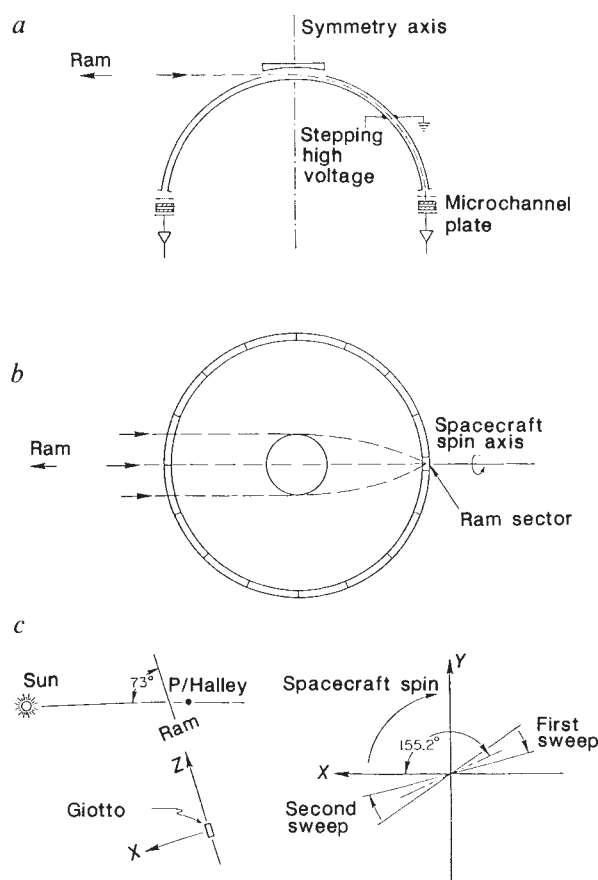


FIG. 1 *a*, Cross-sectional view of the RPA-EESA detector. The open aperture of the detector allows the solar and cometary electromagnetic radiations, as well as dust and neutral gas, to pass through with little chance of colliding with the analyser and producing background. The particle energy E is selected by varying the positive deflection voltage of the inner hemisphere of the electrostatic analyser. *b*, Top view showing how negatively charged particles entering the analyser from a given direction will be focused at the exit. The particles are then detected by a microchannel plate whose 360° collector is divided into 17 sector anodes; 14 of 22.5° width, two of 19.5°, and a 6°-wide ram sector which views particles coming from the direction of spacecraft motion (ram direction). *c*, The EESA planar field of view is oriented parallel to the spacecraft spin axis, which is aligned with the ram direction (direction of propagation of the spacecraft). The second diagram shows the look directions of the detector in the x - y plane when the two ram-sector measurements per spin are made. The z -axis points in the ram direction, and the Sun lies in the x - z plane. The ram-sector response is symmetric in azimuth (defined as angle measured in the 360° plane of EESA field of view) but is a function of both energy and elevation angle (angle from the ram direction)²⁶. Thus, a Maxwellian velocity distribution whose centre is offset from the ram direction will result in spin modulation of the ram-sector count rate.

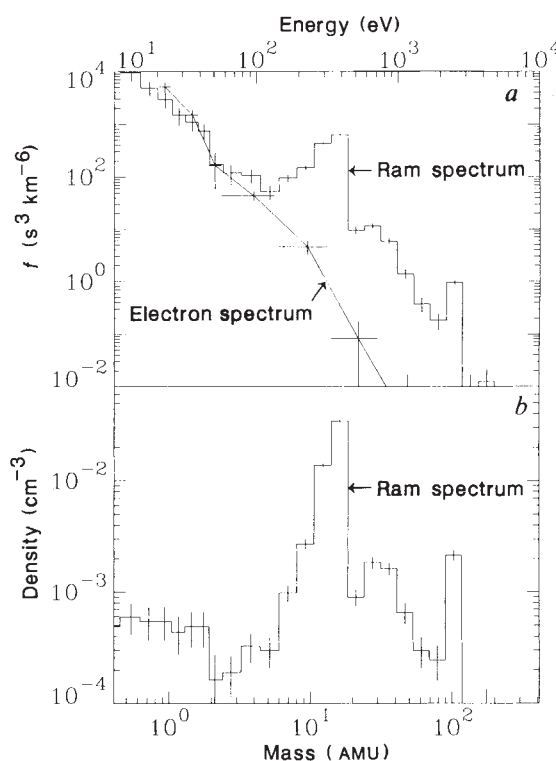
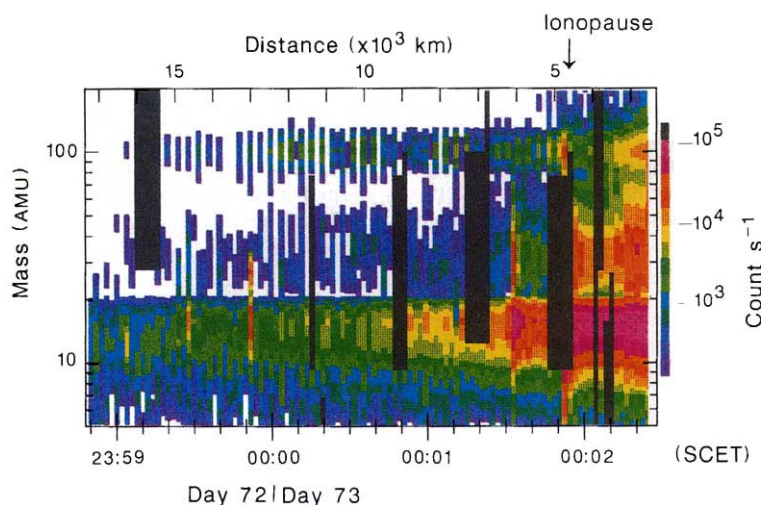


FIG. 2 Energy spectra taken $\sim 5,800$ km from the nucleus. *a*, Spectra in the ram direction and 22° from the ram direction, plotted in units of distribution function (assuming that all particles are electrons). *b*, Three well-defined mass peaks plotted in units of density (assuming the particles are cold ions at rest in the comet frame).

FIG. 3 Colour spectrogram of count rate as a function of mass and time. Distance to closest approach is plotted on the top scale. The colour scale for the count rate is on the right. Spin modulation is clearly apparent outside the ionopause. Lost data are displayed in grey. SCET, Spacecraft Event Time.



density n is observed to vary with distance r as

$$n(r) = c_0 r^{-2} \exp(-c_1 r)$$

where $c_0 = 4.72 \times 10^{13}$ (molecules $\text{cm}^{-3} \text{km}^2$) and $c_1 = 2.51 \times 10^{-5} \text{km}^{-1}$; heavier molecules, such as CO_2 , show a similar variation¹⁹. The r^{-2} variation, produced by the radially outward flow of the gas at an essentially constant velocity, dominates near the comet (the exponential term arises from losses due to ionization). The $\sim r^{-4}$ variation observed for the heavy negative ions inside the ionopause (Fig. 4) is thus inconsistent with this process of formation.

Cometary neutrals and dust may also interact with the plasma clouds that are produced around the spacecraft by the impact and dissociation of cometary dust particles on the spacecraft bumper shield^{20,21}. Negative ions may then be created through sputtering of other incoming dust particles by the plasma cloud, or through electron attachment or charge exchange with incoming neutrals. Most of these negative ions, however, will be produced too close (within a few tens of metres) to the spacecraft for the motional electric field to accelerate them to the velocities required to produce observable spin modulation. We therefore rule out contamination by the spacecraft or the detector as the source of the negative ions.

The observed negative ions are thus of cometary origin. With the mass resolution of the detector ($\Delta M/M \approx 0.26$) we cannot unambiguously identify the species. The first peak, however, includes the highly electronegative species O^- (1.462-eV electron affinity) and/or OH^- (1.828 eV), and the second peak CN^- (3.82 eV). The width of the 17-AMU peak indicates that C^- (1.26 eV) and/or CH^- (1.24 eV) are also present. The ion H^- , however, is not observed above the electron background. Because of the high proportion of water in the comet, O^- and OH^- are expected to be abundant. The ions CN^- , C^- and CH^- are also likely because their neutral and positively charged counterparts have been observed at comet Halley as well as at other comets²². The peak at ~ 100 AMU is unidentified, but heavy, positively charged molecular ions made up of CHO have been detected in significant quantities²³, and many complex CHO molecules are known to have large electron affinities (for example, $\text{C}_6\text{H}_4\text{O}_2$, 1.89 eV; $\text{C}_7\text{H}_5\text{O}$, 1.61 eV). Between ~ 35 and

85 AMU the spectrum is well above background, indicating that negative ions are present in that mass range as well.

Negative-ion production² by electron attachment, dissociative electron attachment and polar photodissociation of neutrals, as well as by sputtering of dust and charge transfer between neutrals and dust, seems to be inadequate by at least one or two orders of magnitude. But fast neutrals, produced by charge exchange and dissociative recombination of fast, positively charged pick-up ions^{1,24}, can undergo charge exchange with slow cometary

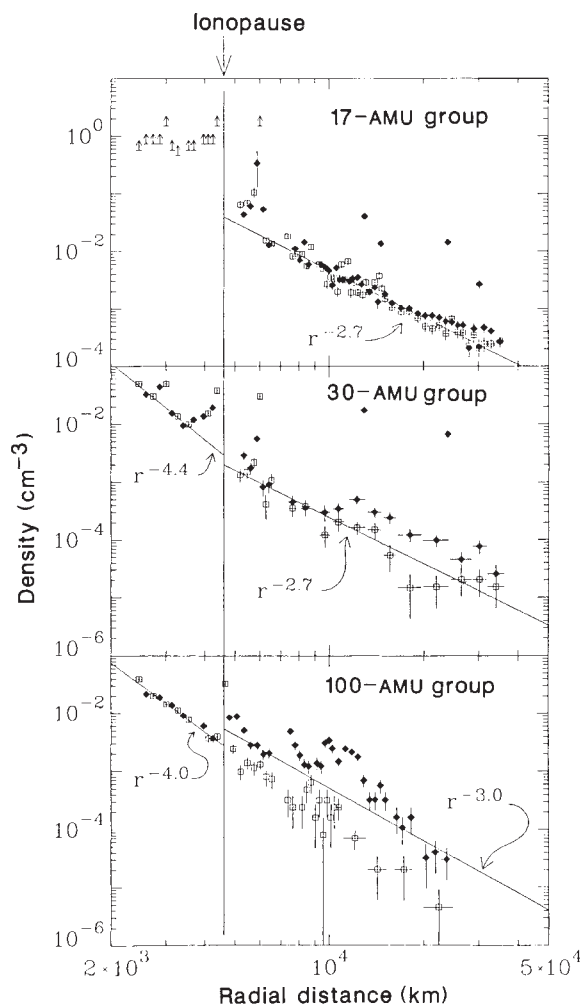
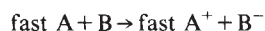


FIG. 4 Radial profile of the density of the three mass groups. The open symbols designate the first sweep and the filled symbols the second sweep (see Fig. 1c), so differences between them indicate spin modulation. Power-law fits to the data are shown both inside and outside the ionopause. No fit is possible for the water group inside the ionopause because the count rate saturated the detector. Spin modulation disappears inside the ionopause boundary.

neutrals in the inner coma to produce both positive and negative ions



A cross-section for this process of $\sim 10^{-17} \text{ cm}^2$ would be required to account for the observed OH^- ion population. The appropriate cross-sections for H_2O and OH are not known, although experiments on charge transfer for hydrogen or deuterium beams in various gases²⁵ obtain some cross-sections close to this value. $\square$

Received 31 August; accepted 27 November 1990.

1. Eviatar, A. & Goldstein, R. *Astrophys. J.* **339**, 545–557 (1989).
2. Wekhof, A. *Moon Planets* **24**, 157–173 (1981).
3. Cheng, A. F. *Astrophys. J.* **242**, 812–827 (1980).
4. Réme, H. et al. *J. phys. Engng* **20**, 721–731 (1987).
5. Réme, H. et al. *Nature* **321**, 349–352 (1986).
6. Réme, H. et al. *Astr. Astrophys.* **187**, 33–38 (1987).
7. Coates, A. J. et al. *Nature* **327**, 489–492 (1987).
8. Anderson, K. A. et al. *Astr. Astrophys.* **187**, 290–292 (1987).
9. d'Uston, C. et al. *Ann. geophys.* **7**, 91–96 (1989).
10. Mazelle, C. et al. *Geophys. Res. Lett.* **16**, 1035–1038 (1989).
11. Neubauer, F. et al. *Nature* **321**, 352–355 (1986).
12. Balsiger, H. et al. *Nature* **321**, 330–334 (1986).
13. Wisenberg, J. & Kockarts, G. *J. geophys. Res.* **85**, 4642–4652 (1980).
14. Kettmann, G. et al. *Ann. Geophys.* **8**, 229–238 (1990).
15. Mitchell, D. et al. *Adv. space Res.* **9**, 35 (1989).
16. Neubauer, F. et al. *J. Phys.* **E20**, 714–720 (1987).
17. Smith, G. P., Lee, L. C., Cosby, P. C., Peterson, J. R. & Moseley, J. T. *J. chem. Phys.* **68**, 3818–3822 (1978).
18. Lee, L. C. & Smith, G. P. *J. chem. Phys.* **70**, 1727–1735 (1979).
19. Krankowsky, D. et al. *Nature* **321**, 326–329 (1986).
20. Grün, E. & Reinhard, R. *Spec. Publ. Eur. Space Ag.* **155**, 7–21 (1981).
21. Arnaudeau, F. et al. *Spec. Publ. Eur. Space Ag.* **224**, 21–37 (1984).
22. Mendis, D. A. et al. *Ann. geophys.* **7**, 99 (1989).
23. Mitchell, D. et al. *Science* **237**, 626–628 (1987).
24. Ip, W.-H. *Astrophys. J.* **353**, 290–296 (1990).
25. Tawara, H. *Atomic Data and Nuclear Data Tables* **22**, 491–525 (1978).
26. Carlson, C. W., Curtis, D. W., Paschmann, G. & Michael, W. *Adv. Space Res.* **2**, 67–70 (1982).

ACKNOWLEDGEMENTS. This work was supported by NASA, the CNES and by the Max-Planck-Gesellschaft zur Förderung der Wissenschaften and the Bundesministerium für Forschung und Technologie.

Phonon suppression of coherence peak in nuclear spin relaxation rate of superconductors

Philip B. Allen* & Dierk Rainer†‡

*Department of Physics, SUNY at Stony Brook, Stony Brook, NY 11794-3800, USA

†Science and Technology Center for Superconductivity, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, USA

THE temperature dependence of the nuclear spin relaxation rate $1/T_1$ peaks sharply (the 'Hebel-Slichter' or 'coherence' peak) just below the transition temperature T_c of a superconductor¹. Because the observation² of this peak definitively confirmed BCS theory³, its absence^{4–6} in the high- T_c oxide superconductors is the best evidence against a BCS picture. Here we show that, to the contrary, an extended form of BCS theory gives a natural explanation. Using the extension by Migdal⁷ and Eliashberg⁸ to retarded interactions (for phonon coupling, the attraction between electrons propagates at the speed of sound, slower than the Fermi velocity of electrons), we predict unexpectedly strong damping effects in all dynamical properties when the temperature T is close to T_c . The origin of the damping (which suppresses the coherence peak) is numerous electron-phonon decay channels open to excitations because of the high T_c itself. This process still works even if another source of attraction beyond electron-phonon coupling causes the high T_c . Thus our observations remove a barrier inhibit-

ing conventional descriptions of high- T_c materials, but by no means force such an interpretation.

Our starting point is the Eliashberg nonlinear integral equations⁹ for the complex, frequency-dependent gap function $\Delta(\omega, T)$. These equations rest on the Fermi-liquid picture, and have been beautifully confirmed by tunnelling experiments¹⁰ near $T = 0$, and by thermodynamic studies¹¹ at temperatures up to T_c . Dynamical properties for $0 < T < T_c$ are harder to compute, and have attracted less theoretical attention. No reliable method exists except numerical solution for particular models. A model is defined by the electron-phonon spectral function $\alpha^2F(\omega)$ and the Coulomb coupling constant μ^* . The coupling constant λ is defined as $2 \int d\omega \alpha^2F(\omega)/\omega$, and measures the strength of the electron-phonon attraction. In BCS theory, T_c is determined by a coupling constant ' $N(0)V$ ' (where $N(0)$ is the density of states at the Fermi level and V is the attractive interaction potential); in a rough approximation, Eliashberg theory uses instead $\lambda - \mu^*$.

These model parameters are known from tunnelling measurements on certain conventional superconductors. Rough guesses can be made for high- T_c superconductors, but it is a matter of conjecture whether these notions apply to the new oxide materials. We take as representative choices three models: (a) indium with coupling constants $\lambda = 0.805$ and $\mu^* = 0.121$ (ref. 12); (b) crystalline alloy $\text{Pb}_{0.9}\text{Bi}_{0.1}$ with $\lambda = 1.66$ and $\mu^* = 0.095$ (ref. 13); and (c) a hypothetical high- T_c material with the $\alpha^2F(\omega)$ spectrum of case (b) and $\lambda = 3.2$ and $\mu^* = 0.1$. Well-converged numerical solutions are found for a series of temperatures $T_n = (1 - 0.036n)T_c$. At the highest temperature, $0.964T_c$, our results for $\text{Pb}_{0.9}\text{Bi}_{0.1}$ agree perfectly with the published results of Vashista and Carbotte [14]. Details will be given in a longer paper.

After finding $\Delta(\omega, T)$, dynamical properties can be calculated. Of central importance is the density of states $N(\omega)$, which is constant for small ω in the normal state, and is given in the superconducting state by

$$N_s(\omega)/N_n(\omega) = \text{Re}[\omega/\sqrt{\omega^2 - \Delta^2(\omega)}] \quad (1)$$

where subscripts s and n refer to the superconducting and normal states. For models (b) and (c), this function is plotted at various temperatures in Fig. 1. For model (a) (indium), the density of

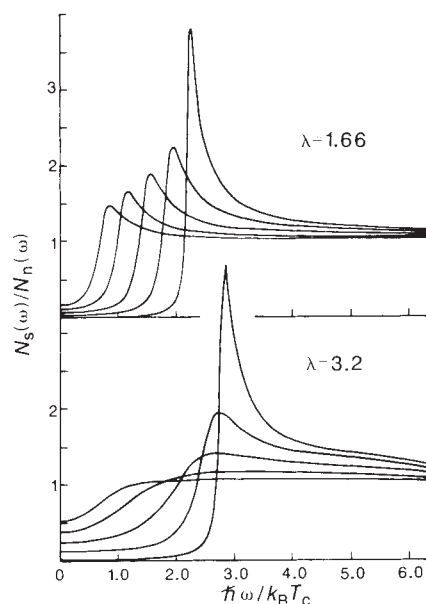


FIG. 1 Density of states (normalized to the normal-state value) versus energy ω calculated from equation (1) for the $\alpha^2F(\omega)$ spectrum of $\text{Pb}_{0.9}\text{Bi}_{0.1}$ at five temperatures: $T/T_c = 0.964, 0.928, 0.856, 0.748$ and 0.568 . (The peak in $N(\omega)$ shifts to the left and broadens with increasing temperature.) The upper panel shows the calculation using the experimental value $\lambda = 1.66$ ('case (b)'); the lower panel is for $\lambda = 3.2$ ('case (c)').

‡Permanent address: Institut für Theoretische Physik III, Universität Bayreuth, D-8580 Bayreuth, Germany.

states curves at various temperatures are sharper than in case (b), and follow quite closely one's expectations based on BCS theory: there are very few states below a well-defined gap edge Δ_1 , at which point there is a near divergence, and a return to the normal-state value at large ω . The gap edge can be defined as the peak of N_s in equation (1), or what is almost the same, by Δ_1 in the equation

$$\Delta_1 + i\Delta_2 \equiv \Delta(\omega = \Delta_1, T) \quad (2)$$

For any temperature $T > 0$ there are in fact a few states below the gap, and the divergence is rounded by inelastic lifetime broadening. (A typical event causing such gap broadening is the evolution of an injected virtual-electron quasiparticle with energy $\omega < \Delta$ into a real excitation, by the absorption of a thermal phonon to provide the missing energy.) Lifetime broadening of the gap was studied using tunnelling by Dynes *et al.*¹⁵, who obtained excellent fits to data by using equation (1) with a real constant for $\Delta(\omega)$ and ω replaced by $\omega - i\Gamma$, where Γ represents the energy uncertainty introduced by a finite lifetime. This prescription worked well for the small broadenings that they measured; the correct prescription is to use equation (1) with the Eliashberg complex gap function $\Delta(\omega)$.

For indium, the broadening is small. However, a factor of two increase in λ going from case (a) to case (b) and then again to case (c) has a large qualitative effect, as shown in Fig. 1. As T increases, the gap for $\text{Pb}_{0.9}\text{Bi}_{0.1}$ (upper panel) decreases in BCS-like fashion, but with large broadening and states below the gap edge. At the highest temperature shown, $0.964T_c$, the gap is only barely defined. Going to case (c) (lower panel), there is a well defined gap only at low temperatures. The ratio $2\Delta/k_B T_c$, which has nearly the BCS value of 3.52 for indium, increases to ~ 5 for case (b) and ~ 6 for case (c). As T increases in case (c), the peak in the density of states does not decrease much in energy. Instead, it broadens into oblivion by $0.9T_c$, with states gradually filling in the previously empty region below the zero-temperature gap. Near T_c , no solutions of equation (2) exist. We have investigated whether this interesting density-of-states variation could be seen in a finite-temperature tunnelling measurement. The answer is that thermal broadening in an S/I/N (superconductor/insulator/normal metal) junction will make the effect hard to distinguish from the more ordinary behaviour predicted for $\lambda = 1.66$. However, in an S/I/S experiment¹⁵, the difference would be detectable.

The BCS prediction for the spin relaxation rate $1/T_1$ is

$$[(1/T_1)_s/(1/T_1)_n]_{\text{BCS}} = 2 \int_{\Delta+\Delta_c}^{\infty} dE (-\partial f/\partial E) \frac{E^2 + \Delta^2}{E^2 - \Delta^2} \quad (3)$$

where f is the Fermi-Dirac statistical occupation factor, $\Delta(T)$ is the BCS gap (a real number, independent of ω) and $\Delta_c(T)$ is a cutoff introduced arbitrarily to prevent the integral diverging. The integral can be approximately evaluated, yielding

$$\begin{aligned} [(1/T_1)_s/(1/T_1)_n]_{\text{BCS}} \\ \approx 2f(\Delta) \left[1 + \frac{\Delta}{k_B T} (1 - f(\Delta)) \ln(2\Delta/\Delta_c) \right] \end{aligned} \quad (4)$$

Hebel and Slichter² interpreted the cutoff as a measure of gap anisotropy. In a dirty superconductor, the gap becomes isotropically averaged. Fibich¹⁶ pointed out that intrinsic lifetime broadening in Eliashberg theory can cut off the divergence. In Eliashberg theory, the formula analogous to equation (3) is

$$\begin{aligned} [(1/T_1)_s/(1/T_1)_n]_{\text{Eliash}} = 2 \int_0^{\infty} d\omega (-\partial f/\partial \omega) \\ \times \{ [\text{Re}(\omega/\sqrt{\omega^2 - \Delta(\omega)^2})]^2 \\ + [\text{Re}(\Delta(\omega)/\sqrt{\omega^2 - \Delta(\omega)^2})]^2 \} \end{aligned} \quad (5)$$

Fibich made approximations that reduced equation (5) to

equation (4), with $\Delta(T)$ and $\Delta_c(T)$ replaced by Δ_1 and Δ_2 of equation (2). For Δ_1 he suggested the BCS gap, and for Δ_2 gave an approximate formula. We have instead done exact numerical evaluations of equation (5). The results are shown in Fig. 2 for all three models. The values for indium agree well with data from ref. 17 for indium with 0.8% Tl intentionally introduced to destroy gap anisotropy. No parameter was adjusted to give this fit. The Fibich formula can be adjusted to give an equally good fit, but requires an unrealistic Debye temperature, 35 K.

Our results for $1/T_1$ for models (b) and (c) are also shown in Fig. 2. Neither of these can be fitted well by the Fibich formula, which is not surprising given that the energy gap is poorly defined in these cases. The complete absence of the coherence peak in case (c) is striking, and accords with experiment. It is, of course, accidental that data for $\text{YBa}_2\text{Cu}_3\text{O}_7$ happen to agree so well with our calculations for model (c) as shown in Fig. 2. The value $\lambda = 3.2$ was chosen based on a reasonable guess of what λ would have to be to give $T_c = 90$ K in a conventional description. However, if such a description applies, we still have little information about the shape of $\alpha^2 F(\omega)$, and the answer for $1/T_1$ depends on the shape, not just on λ .

Apparently, the absence of the coherence peak in experiments is not a clear signal for exotic superconductivity. Dolgov *et al.*¹⁸ have also reached this conclusion. The exact basis for their curve is not clear (they have used "inelastic smearing of the density of states", that is, apparently not $T > 0$ solutions of Eliashberg theory) but the result seems correct. Our conclusions are somewhat related to those of Coffey¹⁹, who used a temperature-dependent pair-breaking scattering to suppress the coherence peak. Inelastic scattering from thermal phonons in our picture provides an intrinsic pair-breaking effect. Our conclusions also accord with those of Kuroda and Varma²⁰, who argue that inelastic scattering of "marginal quasiparticles" from density

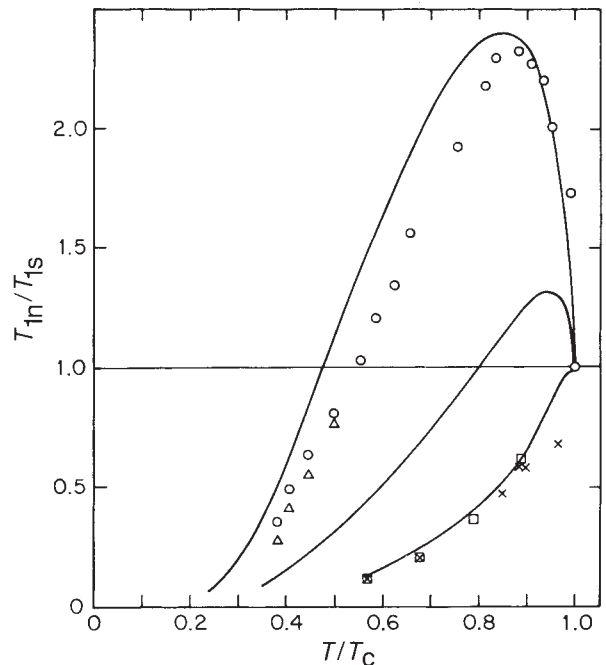


FIG. 2 Nuclear spin relaxation rate $[(1/T_1)_s/(1/T_1)_n]$ versus temperature T/T_c . The data points for indium come from ref. 17: circles and triangles represent data with and without corrections for heating. The solid curves are calculated from equation (5), for indium (upper curve) and for the $\alpha^2 F(\omega)$ spectrum of $\text{Pb}_{0.9}\text{Bi}_{0.1}$ (lower two curves) with $\lambda = 1.66$ and 3.2 respectively. Data for $\text{YBa}_2\text{Cu}_3\text{O}_7$ come from ref. 5: squares are ^{63}Cu resonances from copper atoms in the CuO_2 planes; crosses are ^{17}O resonances from in-plane oxygen atoms. Copper data from ref. 6 agree closely. For all data the value of $1/T_1$ in the normal state was chosen by Korringa-law extrapolation from the measured value just above T_c .

fluctuations of a "marginal Fermi liquid" will also destroy the coherence peak. Their density fluctuations are playing much the same role as phonons in Eliashberg theory. After submitting the first version of this paper we received a preprint from Akis and Carbotte²¹ with Eliashberg calculations of $1/T_1$ in close agreement to ours given here.

Finally, it is appropriate to clarify our position on the nature of high- T_c materials. Our calculations rest on a 'conventional' Fermi-liquid picture. A difficulty in accepting a conventional description is certain anomalous normal-state transport properties, including the temperature-dependent Hall coefficient and the large (and non-Korringa-like) $1/T_1$ of copper nuclei in the CuO_2 planes. To address these issues requires going outside the

boundaries of Eliashberg theory, and possibly even rejecting the underlying Fermi-liquid theory. Our work sheds no light on these issues. Normal-state anomalies are encountered elsewhere in superconductors, namely in the A15 and related d -band compounds, and 'heavy Fermion' metals. In both cases it appears that the Fermi-liquid picture is destroyed at higher temperature but still provides the correct starting point for theories of superconductivity. We feel that it is premature to guess whether this will also hold for the high- T_c materials. What the present paper shows is that the 'conventional' picture, when pushed to the regime of high T_c , has interesting and unexpected anomalies associated with the large inelastic scattering, which resemble experiment—at least in the case of $1/T_1$. □

Received 21 November 1990; accepted 8 January 1991.

1. MacLaughlin, D. E. *Solid St. Phys.* **31**, 1–69 (1977).
2. Hebel, L. C. & Slichter, C. S. *Phys. Rev.* **113**, 1504–1519 (1959).
3. Bardeen, J., Cooper, L. N. & Schrieffer, J. R. *Phys. Rev.* **108**, 1175–1204 (1957).
4. Pennington, C. H. & Slichter, C. P. in *Physical Properties of High Temperature Superconductors* Vol. II (ed. Ginsberg, D. M.) Ch. 5 (World Scientific, Singapore, 1990).
5. Hammel, P. C., Takigawa, M., Heffner, R. H., Fisk, Z. & Ott, K. C. *Phys. Rev. Lett.* **63**, 1992–1995 (1989).
6. Kitaoka, Y. *et al.* in *Strong Correlation and Superconductivity* (eds Fukuyama, H., Maekawa, S. & Malozemoff, A. P.) 262–273 (Springer, Berlin, 1989).
7. Migdal, A. B. *Zh. eksp. teor. Fiz.* **34**, 1438–1446 (1958); *Soviet Phys. JETP* (Engl. trans.) **7**, 996–1001 (1958).
8. Eliashberg, G. M. *Zh. eksp. teor. Fiz.* **38**, 966–976 (1960); *Soviet Phys. JETP* (Engl. trans.) **11**, 696–702 (1960).
9. Scalapino, D. J. in *Superconductivity* Vol. 1 (ed. Parks, R. D.) Ch. 10 (Dekker, New York, 1969).

10. McMillan, W. L. & Rowell, J. M. in *Superconductivity* Vol. 1 (ed. Parks, R. D.) Ch. 11 (Dekker, New York, 1969).
11. Carbotte, J. P. *Rev. Mod. Phys.* (in the press).
12. Dynes, R. C. *Phys. Rev.* **B2**, 644–656 (1970).
13. Dynes, R. C. & Rowell, J. M. *Phys. Rev.* **B11**, 1884–1894 (1975).
14. Vashishta, P. & Carbotte, J. P. *Phys. Rev.* **B7**, 1874–1880 (1973).
15. Dynes, R. C., Narayanamurti, V. & Garno, J. P. *Phys. Rev. Lett.* **41**, 1509–1512 (1978).
16. Fibich, M. *Phys. Rev. Lett.* **14**, 561–564, 621 (erratum) (1965).
17. Williamson, J. D. & MacLaughlin, D. E. *Phys. Rev.* **B8**, 125–132 (1973).
18. Dolgov, O. V., Golubov, A. A. & Koshelev, A. E. *Solid St. Commun.* **72**, 81–83 (1989).
19. Coffey, L. *Phys. Rev. Lett.* **64**, 1071 (1990).
20. Kuroda, Y. & Varma, C. M. *Phys. Rev.* **42**, 8619–8622 (1990).
21. Akis, R. & Carbotte, J. P. *Solid St. Commun.* (in the press).

ACKNOWLEDGEMENTS. We thank B. Mitrovic, J. P. Carbotte and R. C. Dynes for help. Work at Stony Brook was supported in part by NSF. Much of this work was done while the authors were at the Center for Materials Science, Los Alamos National Laboratory, Los Alamos, New Mexico.

Molecular-resolution images of Langmuir–Blodgett films using atomic force microscopy

E. Meyer*, L. Howald*, R. M. Overney*,
H. Heinzelmann*†, J. Frommer*†, H.-J. Güntherodt*,
T. Wagner†, H. Schier† & S. Roth†

* Institut für Physik, Universität Basel, Klingelbergstrasse 82,
CH-4056 Basel, Switzerland

† Max-Planck-Institut für Festkörperforschung, Heisenbergstrasse 1,
D-7000 Stuttgart 80, Germany

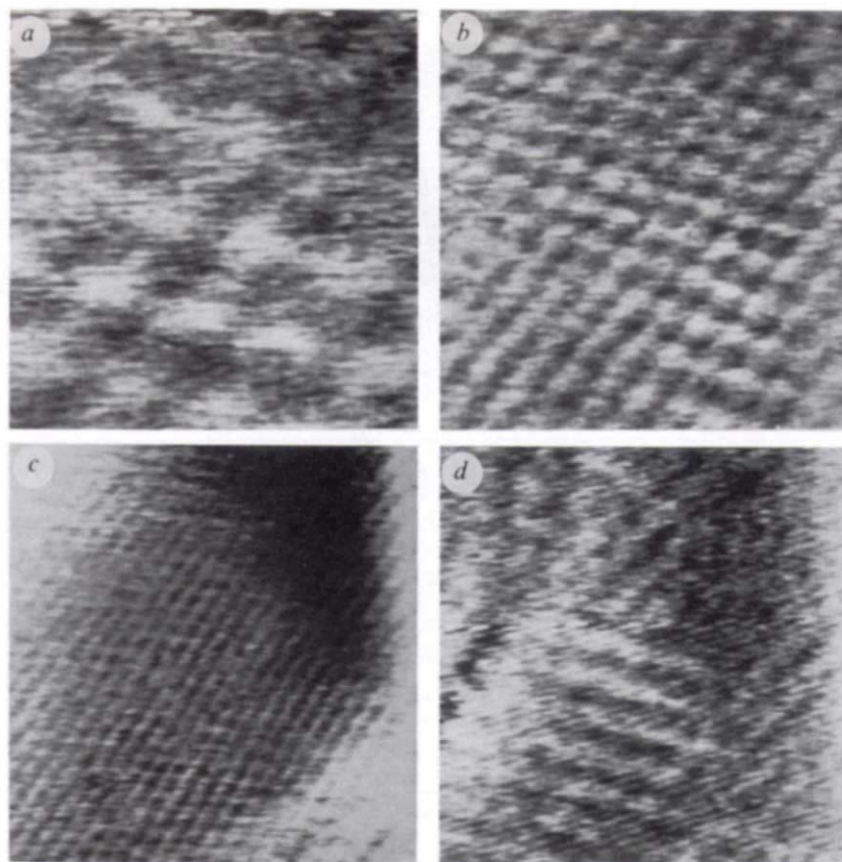
THE ability to prepare thin films of amphiphilic molecules (Langmuir–Blodgett (LB) films) is valuable to many areas of research. In biology they provide models for ideal membranes; the two-dimensional behaviour and structural phase transitions are of fundamental interest in surface physics; and their tribological characteristics suggest potential engineering applications. For determining the structure of these films, the common techniques such as X-ray and neutron scattering are limited to thick (≥ 200 Å) multilayers. Thinner films can be studied by transmission electron microscopy and low-energy electron diffraction^{1,2}, but these electron-beam techniques tend to damage thin films. More recently, the scanning tunnelling microscope³ has provided a non-destructive means of investigating the structures of LB films^{4–6}, but as the films are insulating, the interpretation of such images has been controversial. The atomic force microscope⁷ is not plagued with these ambiguities, as it does not require a conductive sample. Here we present images, with molecular resolution, of LB films of cadmium arachidate deposited on an amorphous silicate substrate. Despite the disorder in the substrate, the films display a periodic structure over large distances (several hundreds of ångströms). This suggests that the adsorbed molecules near the interface are driven to self-assemble primarily, if not solely, by intermolecular forces rather than by dependence on substrate periodicity.

The study of the ordered molecular structure of LB films was one of the first applications of the scanning tunnelling microscope (STM) to organic materials^{4–6}. There has been some controversy over the interpretation of the images recorded on these rather thick films (~ 50 Å) because classical tunnelling theory cannot account for the contrast mechanism observed in these experiments. It has been suggested⁶ that the molecules are pushed aside by the tip until a tip-to-surface distance is reached that can sustain a tunnelling current. In this explanation the STM images are attributed to the graphite substrate rather than to the LB adsorbate. Alternative explanations such as resonant tunnelling have also been proposed⁸. This controversy over the origin of the tunnelling current modulation is avoided by the use of the atomic force microscope (AFM) to image the LB films. The AFM offers the possibility of imaging surfaces on a molecular or atomic scale without the requirement of an electrically conducting sample⁹. In the AFM the probing tip is attached to a cantilever-type spring which is deflected in response to the forces between the probing tip and the sample. The deflections are monitored with sensitive detectors based on electron tunnelling⁷, optical interferometry^{10,11}, laser-beam deflection^{12,13} and capacitance methods^{14,15}. Atomic resolution with the AFM has been achieved on several crystalline layered materials such as graphite, boron nitride, transition-metal dichalcogenides¹⁶ and ionic crystals such as NaCl ¹⁷ and LiF ¹⁸. A polymeric surface stabilized by inter-chain covalent bonding has also been imaged with the AFM¹⁹.

The LB layers in our study are prepared according to standard literature practice²⁰. Silicon substrates (floating zone (111); $p(\text{B})$; $1,000 \Omega \text{ cm}$) were cleaned by the RCA technique²¹, etched in 30% HF and rinsed in Milli-Q water before use. Because the surface of the etched silicon wafer is hydrophobic²² (as determined by contact angle), on the downstroke of the dipping procedure the hydrophobic top surface of the floating monolayer adheres to the wafer. The surface of the coated wafer is now hydrophilic so that on the upstroke, another monolayer adheres with the hydrophilic headgroups pointing inward. This process is repeated once to yield four-layer films of cadmium arachidate. These films are stable over months when stored at moderate humidity and temperatures ($\leq 40^\circ \text{C}$).

† Also affiliated with IBM Almaden Research Laboratory, 650 Harry Road, San Jose, California 95120-6099, USA.

FIG. 1 *a*, AFM image of a four-layer cadmium arachidate film. The scan area is $23 \times 23 \text{ \AA}$ and the atomic corrugation height is $0.3 \text{ \AA}$. *b*, $53 \times 51 \text{ \AA}$ area. The repeat distances between the protrusions are $5.2 \pm 0.3 \text{ \AA}$ and $4.7 \pm 0.3 \text{ \AA}$. *c*, $106 \times 102 \text{ \AA}$ area. Even on this large scale the film still appears well ordered. In addition to the atomic corrugation, one can see height variation due to distortion from the scanning process (right side of image) and due to substrate nonplanarity (upper left and lower right corners). *d*, $53 \times 51 \text{ \AA}$ area. The periodic arrangement of the molecules is interrupted by a defect structure.



The design of our microscope has been described elsewhere²³. The detection method of the instrument is based on electron tunnelling. Because of its high mechanical stability and low thermal drift rates, the microscope works very reliably, achieving height resolution of $\sim 0.01 \text{ \AA}$. The microscope is operated at ambient pressure and is isolated from building vibrations by an antivibration table. For the high-resolution images, the AFM is operated with repulsive forces on the order of 10^{-8} N . The diamond tip is mounted on a microfabricated SiO_2 lever with a spring constant of 0.7 N m^{-1} .

Figure 1*a* shows a $23 \times 23 \text{ \AA}$ image of a four-layer cadmium arachidate film. The molecules are arranged regularly, their long molecular axes orientated orthogonal to the substrate plane and to the scanned plane. When the scan area is enlarged to $53 \times 51 \text{ \AA}$

or $106 \times 102 \text{ \AA}$, as shown in Fig. 1*b* and *c* respectively, regular ordering is still evident. The intermolecular distances are $5.2 \pm 0.3 \text{ \AA}$ and $4.7 \pm 0.3 \text{ \AA}$ along the two lattice vectors. These distances can be deduced from the two-dimensional fast Fourier transform of the data (Fig. 2). This representation of the data allows a facile comparison to the data of an earlier LEED study¹ on cadmium arachidate films. The AFM images of the four-layer film show a different spacing along the two in-plane vectors, indicating an orthorhombic or monoclinic crystal structure. This difference could originate from a distortion of the molecules or the lattice by the scan process. Nevertheless, the packing density (area per molecule) of $24 \pm 2 \text{ \AA}^2$ is in reasonable agreement with literature values for cadmium arachidate films¹.

During the measurements we observed no damage to the films caused by the probing tip. The force of 10^{-8} N that we used is an upper limit; at higher loadings, plastic deformation of the film has been reported²⁴. Over a range of a few thousand ångströms, we observe few discontinuities (defects) in the periodicity of the film (Fig. 1*d*). Since in these images we observe the periodic arrangement of the molecules as well as defect structures, we conclude that the contact region between tip and sample must be as small as a few atoms, a conclusion consistent with previous studies of the effect of contact area on adhesion of LB films²⁵. The successful AFM imaging of LB films, which are stabilized into a planar configuration by intermolecular interactions only, sets a precedent for the use of these high-resolution instruments on fragile molecular systems.

Our study addresses the question of whether the first few monolayers deposited on amorphous substrates are prevented by disorder in the substrate from forming long-range order. On substrates such as etched wafers²², the dynamics of adsorption and interactions at the interface can be completely different to those on the more idealized surface of a single crystal¹. This demonstration that thin LB films can develop long-range order on amorphous surfaces suggests that the driving force for ordering in these films seems to be stronger than expected

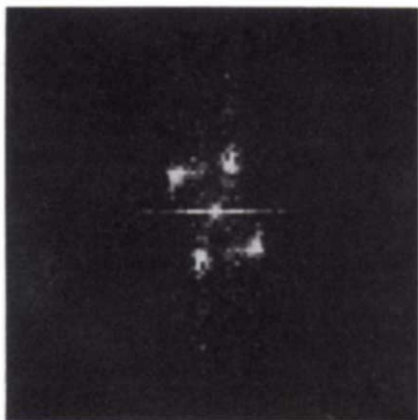


FIG. 2 Two-dimensional fast Fourier transform of Fig. 1*b*. The peaks correspond to the $5.2\text{-\AA}$ and the $4.7\text{-\AA}$ periods. The symmetry indicates an orthorhombic or monoclinic crystal structure.

and to be controlled by interactions between the adsorbed molecules. □

Received 2 October; accepted 17 December 1990.

- Vogel, V. & Wöll, C. *J. chem. Phys.* **84**, 5200–5204 (1986).
- Garoff, S., Deckmann, H. W., Dunsmuir, J., Alvarez, M. S. & Bloch, J. M. *J. Physique* **47**, 701–709 (1986).
- Binnig, G., Rohrer, H. & Gerber, C. *Phys. Rev. Lett.* **49**, 57–61 (1982).
- Smith, D. P. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 969–972 (1987).
- Fuchs, H., Schrepp, W. & Rohrer, H. *Surf. Sci.* **181**, 391–393 (1987).
- Lang, C. A., Hörber, J. K. H., Hänsch, T. W., Heckl, W. M. & Möhwald, H. *J. Vac. Sci. Technol.* **A6**, 368–370 (1989).
- Binnig, G., Quate, C. F. & Gerber, C. *Phys. Rev. Lett.* **56**, 930–933 (1986).
- Mizutani, W., Shigeno, M. & Kajimura, K. *Appl. Phys. Lett.* **56**, 1974–1976 (1990).
- Heinzelmann, H., Meyer, E., Rudin, H., Güntherodt, H.-J. in *Scanning Tunneling Microscopy and Related Methods* (eds Behm, R. J. *et al.*) 443–467 (Kluwer Academic, Netherlands, 1990).
- McClelland, G. M., Erlandsson, R. & Chiang, S. *Rev. Progr. Qual. non-destr. Eval.* **6**, 1307–1315 (Plenum, New York, 1987).

- Rugar, D., Mamin, H. J., Erlandsson, R., Stern, J. E. & Terris, B. D. *Rev. Sci. Instrum.* **59**, 2337–2340 (1988).
- Meyer, G. & Amer, N. M. *Appl. Phys. Lett.* **53**, 1045–1047 (1988).
- Alexander, S. *et al. J. appl. Phys.* **65**, 164–167 (1989).
- Erlandsson, R., McClelland, G. M., Mate, C. M. & Chiang, S. *J. Vac. Sci. Technol.* **A6**, 266–270 (1988).
- Göddenhenrich, T., Hartmann, U., Lemke, H. & Heiden, C. *J. Vac. Sci. Technol.* **8**, 383–387 (1990).
- Meyer, E. *et al. Europhys. Lett.* **9**, 695–700 (1989).
- Meyer, G. & Amer, N. M. *Appl. Phys. Lett.* **56**, 2100–2101 (1990).
- Meyer, E., Heinzelmann, H., Rudin, H. & Güntherodt, H.-J. *Z. Phys. B* **79**, 3–4 (1990).
- Marti, O. *et al. Science* **239**, 50–52 (1988).
- Schreck, M. *et al. Thin Solid Films* **175**, 95–101 (1989).
- Kern, W. *RCA Engineer.* **28**, 99–105 (1983).
- Trucks, G. W., Raghavachari, K., Higashi, G. S. & Chabal, Y. J. *Phys. Rev. Lett.* **65**, 504–507 (1990).
- Meyer, E. *et al. J. Microsc.* **152**, 269–280 (1988).
- Blackman, G. S., Mate, C. M. & Philpott, M. R. *Phys. Rev. Lett.* **65**, 2270–2273 (1990).
- Burnham, N. A., Dominguez, D. D., Mowery, R. L. & Colton, R. J. *Phys. Rev. Lett.* **64**, 1931–1934 (1990).

ACKNOWLEDGEMENTS. We thank R. Steiger, L. Eng and H. Fuchs for stimulating discussions. Technical support from H.-R. Hidber, R. Schnyder and A. Tonin is gratefully acknowledged. This work was supported by Swiss National Science Foundation, the Kommission zur Förderung der wissenschaftlichen Forschung, the Deutsche Forschungsgemeinschaft (SFB 329) and Esprit Basic Research Action OLDS.

Volume transition in a gel driven by hydrogen bonding

Franck Ilmain*, Toyochi Tanaka & Etsuo Kokufuta*

Department of Physics and Center for Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

*Permanent addresses: Laboratoires d'Ultrasons et de Dynamique des Fluides Complexes, Université Louis Pasteur, Strasbourg, France (F.I.); Institute of Applied Biochemistry, University of Tsukuba, Tsukuba-shi, Ibaraki 305, Japan (E.K.).

INTERACTIONS between macromolecules fall into four categories: ionic, hydrophobic, van der Waals and hydrogen bonding. Phase transitions in polymer gels provide a means of studying these interactions. Many gels will undergo reversible, discontinuous volume changes in response to changes in, for example, temperature, gel composition or light irradiation^{1–5}. These transitions result from the competition between repulsive intermolecular forces, usually electrostatic in nature, that act to expand the polymer network, and an attractive force that acts to shrink it. Volume transitions in gels have been observed that are driven by all of the above-mentioned forces except hydrogen bonding (ref 6–10; T.T. *et al.*, unpublished data; H. Inomata *et al.*, personal communication). Here we report on a phase transition in an interpenetrating polymer network of poly(acrylamide) and poly(acrylic acid) that completes this picture—it is controlled by cooperative 'zipping' interactions between the molecules which result from hydrogen bonding. Cooperativity is an essential feature of the interactions, in that independent hydrogen bonds would not provide a sufficient driving force for the transition. A further novel characteristic of this phase transition is that the swelling (in water) is induced by an increase rather than a decrease in temperature.

It is known^{11–16} that poly(acrylamide) and poly(acrylic acid) form polycomplexes in solution through hydrogen bonding. The

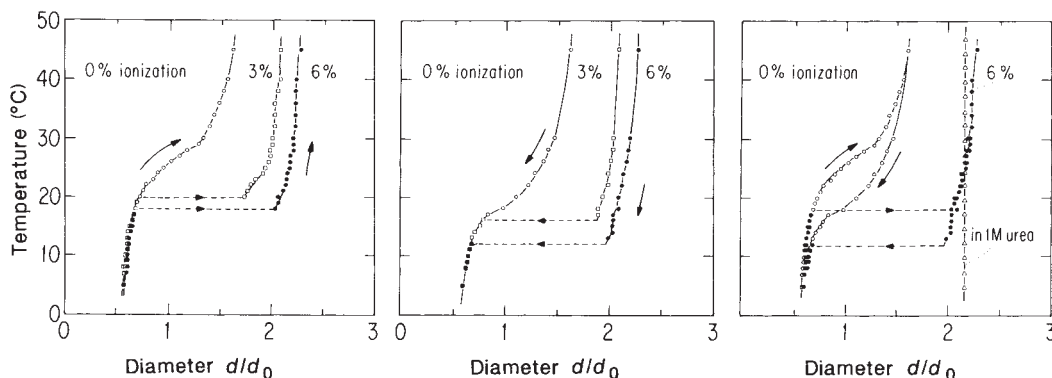
interpenetrating polymer network formed from these two polymers has been used for drug delivery systems, although volume transitions have not been observed¹⁷. To prepare such a network, we first prepared the acrylamide gel: 5 g of acrylamide and 0.133 g of *N,N'*-methylenebisacrylamide (BIS, crosslinker) were dissolved in 100 ml of water, to which 120 µl of tetramethylethylenediamine (accelerator) and 40 mg of ammonium persulfate (APS, initiator) were added⁸. The solution was inserted into capillaries of 0.34 mm inner diameter for gelation. The gels were extracted from the capillaries, washed with water, dried, and then inserted into capillaries of the same diameter. The dried gels were allowed to swell in a solution of 5 g distilled acrylic acid, 0.133 g BIS, 40 mg APS, and varying amounts of NaOH (which is necessary to ionize a fraction of the acrylic acid monomers; ionization was varied from 0 to 9 mol%) in 100 ml of water. Extensive polymerization took place at 60 °C.

Gel diameters as a function of temperature (± 0.01 °C) were measured using a microscope. Figure 1a shows the swelling curves obtained as temperature was increased. The non-ionized gel undergoes a continuous transition at about 20 °C. The ionic gels undergo a discontinuous transition, often with the swollen and collapsed phases coexisting over a wide range of temperature. The horizontal lines in Fig. 1a denote the temperatures at which the collapsed phase disappears. The transition temperature decreases with increasing ionic content.

Figure 1b shows the swelling curves obtained as temperature was decreased. For the ionized gels, the transition temperature varies in different regions of the gel; the transition in Fig. 1b corresponds to the temperature at which a collapsed region first appears. The temperatures for collapse are lower than the temperatures for swelling (Fig. 1c). Hysteresis is more pronounced as the ionic content of the network is increased. The large hysteresis and relatively slow kinetics reflect the cooperative nature of the polycomplexation process.

To confirm the involvement of hydrogen bonding, we studied the effect on the phase transition of adding urea, which is known

FIG. 1 Equilibrium diameter of the interpenetrating polymer networks of poly(acrylic acid) and poly(acrylamide) in water as a function of temperature. a, Increasing temperature. Original diameter $d_0 = 0.34$ mm. The percentage ionization refers to the acrylic acid groups. The transition is continuous for non-ionic gels but discontinuous in gels with ionization above 3%. b, Decreasing temperature. c, Superposition of two sets of data from a and b, and results for non-ionic polymer network in 1M urea solution (open triangles).



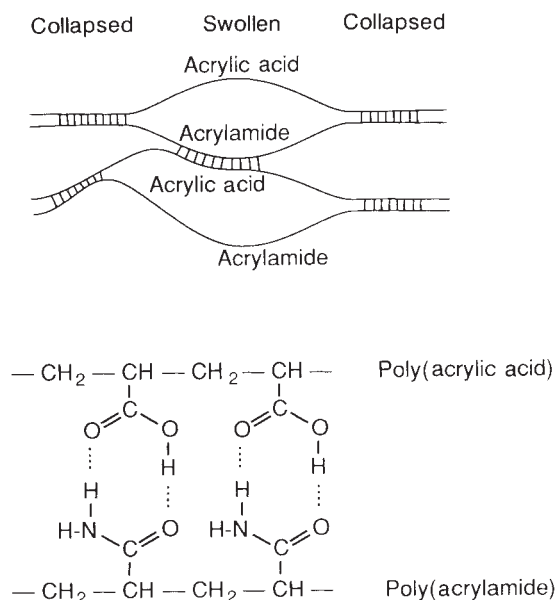


FIG. 2 Possible molecular configuration accompanying the phase transition of the interpenetrating polymer network of poly(acrylamide) and poly(acrylic acid).

to disrupt hydrogen bonds. In the presence of 1M urea, a phase transition did not occur—the gels remained swollen over the entire temperature range (Fig. 1c). We checked that urea did not alter the phase behaviour of an acrylic acid/acrylamide copolymer gel, an acrylic acid/*N*-isopropylacrylamide copolymer gel, an acrylamide gel or an acrylic acid gel: it affected only the poly(acrylamide)/poly(acrylic acid) interpenetrating polymer network. Furthermore, in this system the only attractive force present to induce a transition is hydrogen bonding. These observations suggest that hydrogen bonding must indeed be responsible for the collapse of the gels. A possible molecular configuration that may accompany the gel phase transition is shown in Fig. 2.

It is now possible to classify all gel phase transitions in terms of the four biologically relevant intermolecular forces (Fig. 3), each of which may independently be responsible for a discontinuous volume transition in polymer gels. □

Received 27 June; accepted 12 December 1990.

1. Tanaka, T. *Phys. Rev. Lett.* **40**, 820–823 (1978).
2. Dusek, K. & Patterson, D. *J. Polymer Sci.* **6**, (A-2), 1209–1216 (1968).
3. Tanaka, T. *et al. Phys. Rev. Lett.* **45**, 1636–1639 (1980).
4. Ilavsky, M. *Macromolecules* **15**, 782–788 (1982).
5. Jozuki, A. & Tanaka, T. *Nature* **346**, 345–347 (1990).
6. Hirokawa, Y. & Tanaka, T. *J. chem. Phys.* **81**, 6379–6380 (1984).
7. Otake, K. *et al. J. chem. Phys.* **91**, 1345–1350 (1990).
8. Hirotsu, S., Hirokawa, Y. & Tanaka, T. *J. chem. Phys.* **87**, 1392–1395 (1987).
9. Wakayama, M., Uemoto, S., Okui, T. & Sakai, T. *Polymer Prep. Japan* **38**, 3892–3895 (1989).
10. Myoga, A. & Katayama, S. *Polymer Prep. Japan* **36**, 2852–2853 (1987).
11. Osada, Y., *J. Polymer Sci. Polymer Chem. Ed.* **17**, 3485–3498 (1979).
12. Painter, P. C., Graf, J. & Coleman, M. M. *J. chem. Phys.* **92**, 6166–6174 (1979).
13. Baranovsky, Yu. V., Litmanovich, A. A., Papisov, I. M. & Kabanov, V. A. *Europ. Polymer J.* **17**, 969–979 (1981).
14. Eustace, D. J., Siano, D. B. & Drake, E. N. *J. Appl. Polymer Sci.* **35**, 707–716 (1988).
15. Tsuchida, E. & Abe, K. *Interaction between Macromolecules in Solution* (Springer, Berlin, 1982).
16. Abe, K. & Koide, M. *Macromolecules* **10**, 1259–1265 (1977).
17. Imano, T., Bae, Y. H., Jacobs, H. & Kim, S. W. *J. Controlled Release* **11**, 255–265 (1990).
18. Franks, F. (ed.), *Water, A Comprehensive Treatise* (Plenum, New York, 1973).

ACKNOWLEDGEMENTS. We thank M. Tokita and M. Annaka for valuable discussions. The work was supported by NSF. The stay of F.I. at MIT was supported by Orkem.

Long-range transport of volcanic ash to the Greenland ice sheet

Michael Ram & Robert I. Gayley

Department of Physics, Fronczak Hall, The University at Buffalo, State University of New York, Amherst, New York 14260, USA

BETZER *et al.*¹ reported the surprising discovery of 'giant' (>75 μm) mineral particles transported more than 10,000 km from their source. We have found volcanic ash containing glass shards as large as 300 μm in a section of Wisconsinian ice from the Dye 3 core in Greenland, which is similarly unexpected and raises interesting questions concerning the long-range aeolian transport of particulate matter. As the volcanic ash correlates well with similar ash in deep-sea sediment cores from the Atlantic, it also allows us to date the ice with some confidence.

The aggregates we found in the ice section range in size from a fraction of a millimetre to 3 mm and are scattered over a 78-cm-long section of ice core (Fig. 1). We examined two of the larger aggregates and found that they consisted mostly of highly silicic shards of clear glass, many of them >100 μm long. We also examined particles from one of the fine, parallel cloudy bands and found that they consist of similar volcanic ash particles. All the cloudy ash bands are parallel and inclined at an angle of 70° relative to the core axis, and several of them are very faint.

We retrieved one of the larger (3-mm) aggregates located 9 cm from the bottom of the section (Fig. 1) and examined it by scanning electron microscope (SEM). Figure 2a shows a representative group of particles from the aggregate. Most particles

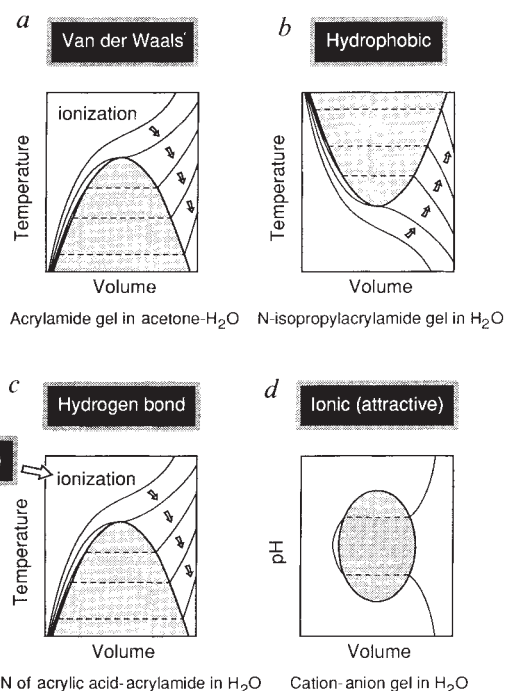


FIG. 3 Gel volume phase transitions induced by four types of intermolecular forces. *a*, The van der Waals interaction causes phase transitions in hydrophilic gels in a mixed solvent, such as an acrylamide gel in an acetone–water mixture. A non-polar solvent is needed to decrease the dielectric constant of the solvent. *b*, Hydrophobic gels, such as *N*-isopropylacrylamide gel, undergo phase transitions in pure water, from a swollen state at low temperatures to a collapsed state at high temperatures. *c*, Gels with cooperative hydrogen bonding, such as an interpenetrating polymer network (IPN) of acrylic acid/acrylamide gel, undergo phase transitions in pure water (the swollen state is the high-temperature state). The repulsive ionic interaction determines the transition temperature and the volume change at the transition³. *d*, The attractive ionic interaction is responsible for pH-driven phase transitions, such as that in acrylamide–sodium acrylate/methacrylamidopropyltrimethylammonium chloride gels¹⁰.

are thin platy glass shards with sharp edges; all are fresh with no signs of corrosion. Numerous adhering small shards can be seen. All signs indicate that the ash was the product of an explosive volcanic eruption². Examination under a light microscope with crossed polarizers revealed that most of the particles were clear glass shards, and we could not identify any lithic bedrock material, nor did we find evidence of bedrock material in the cloudy bands. This is in contrast to earlier reports^{3,4} that the dust layer at this level was granitic bedrock material. As well as the clear glass shards in the aggregate, we found some light- and dark brown particles and a few black ones. X-ray analysis revealed (Table 1) that these are also of volcanic origin. We also examined a 2.5-mm aggregate retrieved 16 cm from the top of the section (Fig. 1) and found that most aggregate particles were clear glass shards, very similar to those of the first aggregate (Fig. 2b).

Particles from the second aggregate were carbon coated and an energy-dispersive X-ray analysis was conducted on some of the larger shards and coloured particles (Table 1). Except for magnetite, where only one particle was measured, the weight percentages are averages over several particles and only major oxides are listed. The quoted error corresponds to 1 s.d. We could not carry out ZAF corrections but measurements were calibrated relative to a KN-18 glass standard⁵. Most of the particles analysed included a fraction of a per cent of Cl, which is a commonly occurring volcanic volatile⁶. In Fig. 3, we show typical particles from the four observed groups. Optical microscopy of particles from a cloudy band showed that they were similar to those found in the two aggregates and that most were clear glass shards.

As there are no active volcanoes in Greenland and as we found no lithic components, we believe that the ash is volcanic fallout carried to the ice sheet by aeolian transport. Because of their size, the large aggregates could not have been deposited in the ice in aggregate form, but were probably formed *in situ*

when ice melting occurred in the summer or was induced by the presence of the ash. When the melt water refroze, it brought together ash particles and formed the visible aggregates.

The presence of numerous giant shards, many larger than 100 μm and some as large as 300 μm , is unexpected because it is generally believed⁷ that such large particles cannot be carried over large distances by aeolian transport. The closest source, Iceland, is 1,000 km away which seems prohibitively far. Our findings seem to require revision of current ideas about aeolian transport. Revision has also been suggested by Betzer *et al.*¹, who found silica particles larger than 100 μm more than 10,000 km from their desert source. In our case, however, it is possible that a powerful explosive volcanic eruption could have lifted particles into the stratosphere, considerably increasing their range of transport. Also, it is likely that increased storminess and strong convection during the Wisconsin could have increased the range of ash transport far beyond that expected during non-glacial times.

One of the problems with the Wisconsinian ice from Dye 3 is that it has been difficult to date precisely^{3,8}. By comparing the oxygen-isotope record of Dye 3 with that of Camp Century, Greenland (77° N, 60° W, 1,885 m above sea level) ice core and the latter with the isotope record from a deep-ocean sediment core from the south Indian Ocean, Dansgaard *et al.*³ obtained an estimate for the age of the Dye 3 Wisconsinian ice that places the ice we studied at ~70,000 yr BP (before present). From measurements of volcanic trace metals in Camp Century ice, Herron⁹ has conjectured a possible high level of volcanic activity during the period 69,000–79,000 yr BP. He suggests that this volcanic signal could correspond to the Z₂ ash layer observed

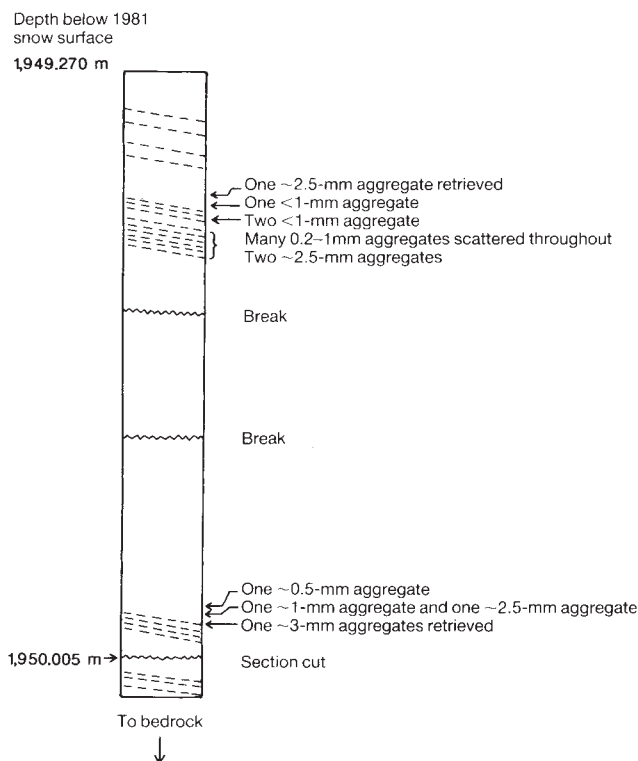


FIG. 1 Section of Wisconsinian ice from Dye 3, Greenland (65° N, 44° W, 2,480 m above sea level) showing position of cloudy bands (dotted lines) and aggregates. Using the chronology of Dansgaard *et al.*³, we estimate that this 78-cm section spans ~400 years. This section is from 1,949 m below the 1981 surface (88 m above bedrock).

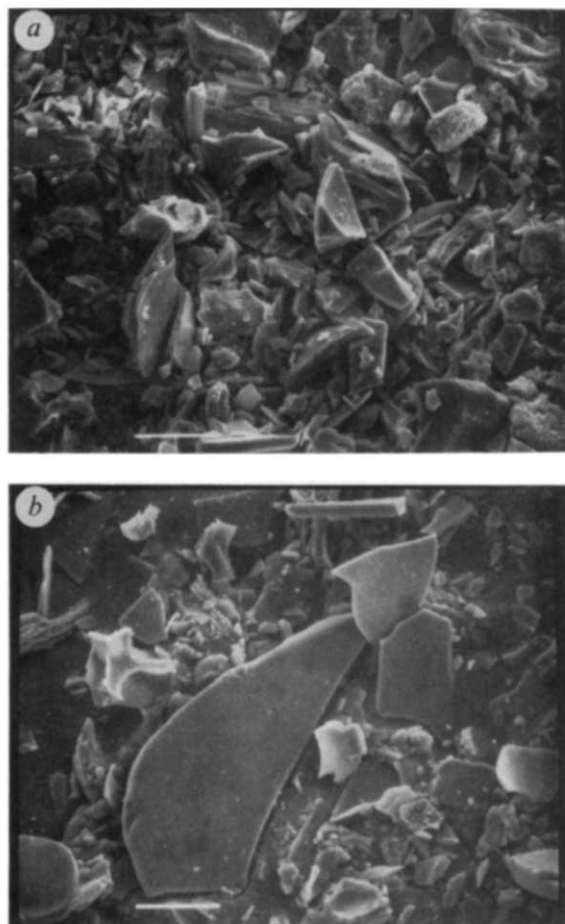


FIG. 2 SEM micrographs (300 $\times$ magnification) of particles from the a, 3-mm aggregate and b, 2.5-mm aggregate. The reference lines are 50 μm long.

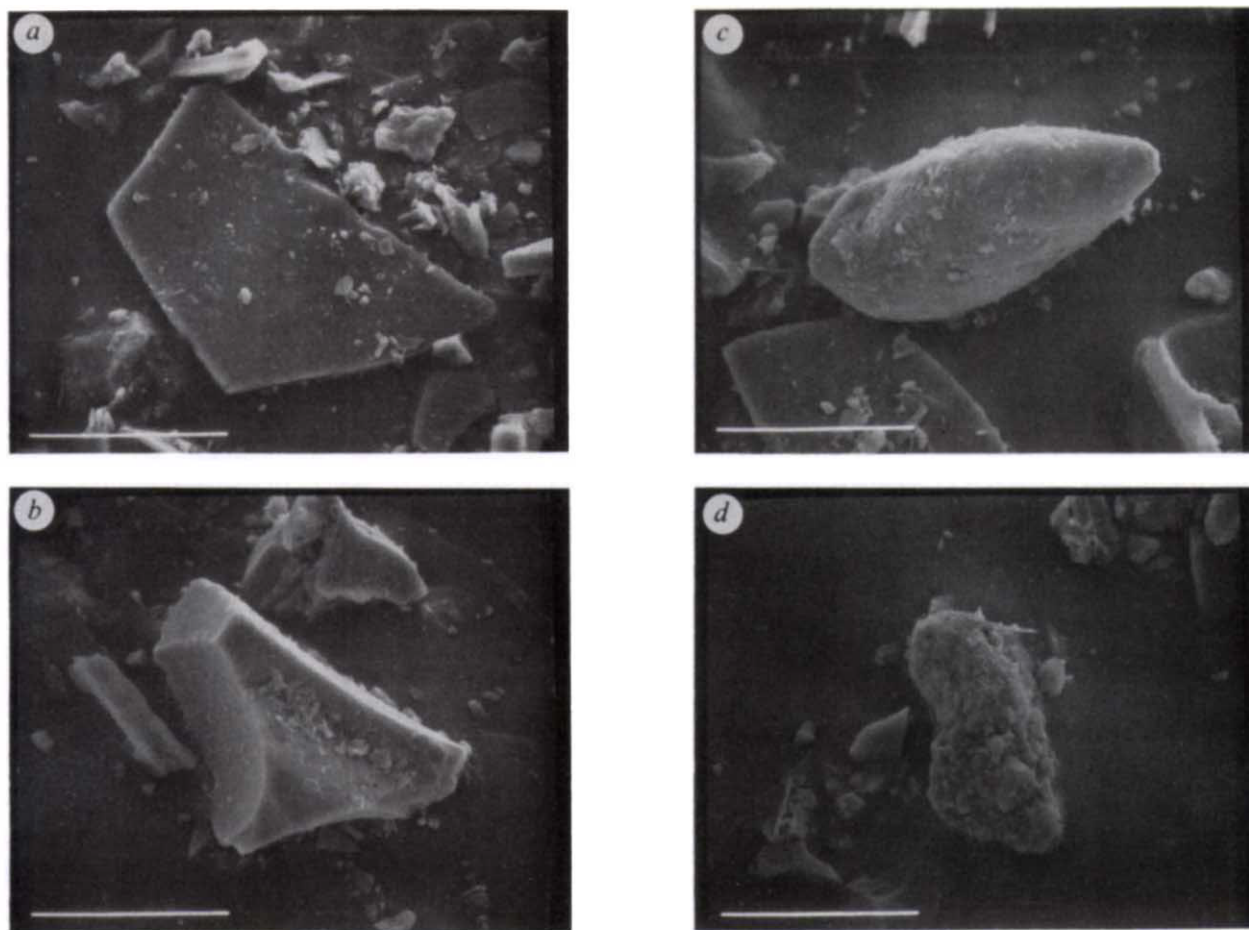


FIG. 3 SEM micrographs (700 $\times$ magnification) of representative particles from each of the four groups in Table 1. *a-d* correspond to the particle groups in columns 1–4 of Table 1, respectively. The central large particle

in each of the micrographs was the one analysed. Reference lines are 50 μ m long.

by Ruddiman and Glover¹⁰ in sediment cores from the north Atlantic. If the Dansgaard *et al.* estimate for the age of our ice is correct, it is reasonable to identify the ash we found with ash zone Z_2 which Ruddiman and Glover dated at $64,700 \pm 3,500$ yr BP. The age estimates are compatible and the Icelandic and Jan Mayen volcanoes, which are the most probable sources of Z_2 ash, are also by far the closest possible sources to Dye 3. Sigurdsson¹¹ has analysed the Z_2 ash and its composition is remarkably similar to ours. We cannot entirely exclude the great Toba, Indonesia, eruptions 75,000 yr BP¹² (which also produced highly silicic ash) as the source of the Dye 3 ash, but the coarseness of our ash and the great distance involved makes Toba an unlikely candidate. Neither can we completely exclude

the Alaskan volcanoes that are known¹³ to have produced highly silicic ash similar to the one we have observed, but their great distance from Greenland make them improbable candidates for the source of the Dye 3 ash.

Ruddiman and Glover¹⁰ measured several secondary volcanic ash peaks in the Z_2 layer. Although this could indicate multiple eruptions, they believe that it is probably due to ice rafting. As ice rafting was of no importance in ash deposition in Greenland, the simplest mechanism that can explain the spreading out of our ash over a 78-cm-long section of ice is multiple volcanic eruptions. If that is the case, and if the origin of the Dye 3 and Z_2 ash are the same, this requires reexamination of the vertical mixing calculations of Ruddiman and Glover¹⁰ who assumed that the Z_2 ash was the result of a single eruption. Examination of the original log of the Dye 3 core shows that the aggregates and cloudy bands we observed were recorded and that no similar events were seen in other parts of the Wisconsinian ice.

The steeply inclined nature of the parallel cloudy ash bands cannot be explained by borehole inclination at drill time, which was at most 6° from the vertical¹⁴. The tilt of the cloudy bands can only be accounted for by deformations that occurred¹⁵ as the ice moved over the irregular bedrock in the Dye 3 area¹⁶. The inclination of the bands could be critical to reassessment of the current depth–age relationships for the core, and the exact nature of the deformation should be revealed by crystal fabric analysis, which was not done on cores deeper than 1,800 m.

The only volcanic horizons that have been reported in Greenland ice cores are high-acidity layers^{17–21} and concentration peaks in volcanic trace metals⁹. The volcanic acidity horizons

TABLE 1 Composition (wt %) of particles in second aggregate

Morphology:	Shards	Blocky	Roundish	—
Colour:	Clear	Light to dark brown	Dark brown to black	Black
Classification:	Rhyolitic	Basaltic	Dacitic	Magnetite
Na ₂ O	5.9 $\pm$ 1.3	5.0 $\pm$ 2.8	5.8 $\pm$ 2.4	—
MgO	—	1.3 $\pm$ 0.3	0.4 $\pm$ 0.2	—
Al ₂ O ₃	11.7 $\pm$ 0.8	8.2 $\pm$ 1.4	9.4 $\pm$ 1.3	0.7
SiO ₂	77.3 $\pm$ 1.6	54.8 $\pm$ 1.7	62.1 $\pm$ 2.2	9.1
K ₂ O	2.9 $\pm$ 0.5	0.8 $\pm$ 0.2	1.5 $\pm$ 0.2	0.3
CaO	—	11.9 $\pm$ 1.0	7.6 $\pm$ 1.6	0.6
TiO ₂	0.1 $\pm$ 0.1	5.3 $\pm$ 0.8	2.9 $\pm$ 1.1	3.3
FeO	2.0 $\pm$ 0.6	12.5 $\pm$ 1.5	10.4 $\pm$ 3.2	86.1

in the ice have been used¹⁷⁻²¹ as markers for dating of Holocene ice cores. No similar volcanic acidity markers exist for Wisconsinian ice, which is alkaline⁴. The Dye 3 ash we found is an important marker and its identification with the Z₂ Ruddiman and Glover ash layer confirms the estimate³ of 70,000 yr BP for the age of the ice. Such confirmation is important because, as pointed by Dansgaard and Oeschger⁸, the Dansgaard *et al.* estimate involved personal judgement, which can be questioned. □

Received 11 September; accepted 28 November 1990.

1. Betzer, P. R. *et al.* *Nature* **336**, 568-571 (1988).
2. Heiken, G. & Wohletz, K. *Volcanic Ash* (University of California Press, Berkeley, 1985).
3. Dansgaard, W., Clausen, H. B., Gundestrup, N., Johnsen, S. J. & Rygner, C. *Am. geophys. Un. Monogr.* **33**, 71-76 (1985).
4. Hammer, C. U. *et al.* *Am. geophys. Un. Monogr.* **33**, 90-94 (1985).
5. Devine, J. D., Sigurdsson, H., Davis, A. N. & Self, S. *J. geophys. Res.* **89**, 6309-6325 (1984).
6. Palais, J. M. & Sigurdsson, H. *Am. geophys. Un.* **52**, 31-53 (1989).
7. Twomey, S. *Atmospheric Aerosols* (Elsevier, New York, 1977).
8. Dansgaard, W. & Oeschger, H. *The Environmental Record in Glaciers and Ice Sheets, Dahlem Workshop Reports* (eds Oeschger, H. & Langway, C. C. Jr) 287-317 (Wiley, New York, 1989).
9. Herron, M. M. thesis, Univ. Buffalo (1980).
10. Ruddiman, W. F. & Glover, L. K. *Geol. Soc. Am. Bull.* **83**, 2817-2836 (1972).
11. Sigurdsson, H. *Eldur Er I Nordri*, 119-127 (Sogufelag, Reykjavik, 1982).
12. Nirkovitch, D., Shackleton, N. J., Abdel-Monem, A. A., Obradovich, J. D. & Izett, G. *Nature* **276**, 574-577 (1978).
13. Westgate, J. *Geophys. Res. Lett.* **15**, 376-379 (1988).
14. Hansen, B. L. & Gundestrup, N. S. *J. Glaciol.* **34**, 178-182 (1988).
15. Whillans, I. M. & Jezek, K. C. *J. geophys. Res.* **92**, 485-493 (1987).
16. Reeh, N. *The Environmental Record in Glaciers and Ice Sheets, Dahlem Workshop Reports* (eds Oeschger, H. & Langway, C. C. Jr) 141-159 (Wiley, New York, 1989).
17. Hammer, C. U. *Nature* **270**, 482-486 (1977).
18. Hammer, C. U. *J. Glaciol.* **25**, 359-372 (1980).
19. Hammer, C. U., Clausen, H. B. & Dansgaard, W. *Nature* **288**, 230-235 (1980).
20. Hammer, C. U., Clausen, H. B. & Dansgaard, W. *J. Volcan. geotherm. Res.* **11**, 3-10 (1981).
21. Hammer, C. U. *Jökull* **34**, 51-65 (1984).

ACKNOWLEDGEMENTS. We thank R. Jacobi and P. Weber for assistance with the analyses, H. Sigurdsson for providing the glass standard, C. C. Langway Jr for making available the Dye 3 ice, L. Burckle, A. Gow, H. Sigurdsson and J. Westgate for reviews and A. Gow for discussions. This work was supported in part by the NSF.

Coherence between seasonal cycles of dimethyl sulphide, methanesulphonate and sulphate in marine air

G. P. Ayers*, J. P. Ivey† & R. W. Gillett*

* Division of Atmospheric Research, CSIRO, Private Bag 1, Mordialloc 3195, Australia

† Australian Government Analytical Laboratories, Channel Highway, Kingston 7051, Australia

THE effect of cloud condensation nuclei (CCN) in controlling cloud albedo and hence global climate has prompted questions about the factors that control CCN populations. Charlson *et al.*¹ suggested that marine phytoplankton might control CCN populations. The effect would be strongest over the oceans, where low stratiform clouds are common, and CCN are composed mostly of ammonium sulphate believed to be produced by atmospheric oxidation of gaseous dimethyl sulphide (DMS) emitted by phytoplankton. Here we present twenty months of data from a clean marine site at 40° S, which confirm the connection between atmospheric DMS and aerosol sulphur species that is central to the hypothesis of Charlson *et al.*¹. The relationships between DMS, aerosol sulphate and CCN are nonlinear, implying that at this site there would be significant nonlinearities in the climate feedback mechanism. In addition, our data on atmospheric sulphur species show a strong seasonal cycle, indicating that it should be possible to look for large, natural, seasonal variations in cloud albedo as a rigorous test of the Charlson *et al.* hypothesis.

Here we present a time series of concurrent observations aimed at assessing the coherence between gaseous DMS and its

presumed aerosol oxidation products, methanesulphonate (MSA) and non-sea-salt (n.s.s.) sulphate. In previous concurrent measurements of these three species during short-duration field expeditions²⁻⁴, covariance between DMS and the aerosol sulphur components was not obvious. Our aim was to take advantage of the large systematic seasonal variations expected in DMS emissions at mid-to-high latitudes⁵ to investigate coherence over seasonal timescales. Our measurements were made at Cape Grim, Tasmania (40° 41' S, 144° 41' E), an ideal site for this investigation as the unpolluted marine air masses of the 'roaring forties' sampled at Cape Grim typify those on which the hypothesis of Charlson *et al.* is based.

Weekly samples of aerosol were collected by filtration between November 1988 and May 1990, covering two summer peaks in DMS emissions. Sampling was carried out at the Cape Grim observatory, 94 m above sea level, only during 'baseline' conditions—periods defined by wind direction from the ocean (180°-290°) and condensation nucleus concentration <600 cm⁻³. These conditions occur ~30% of the time at Cape Grim.

The high atmospheric load of sea salt at windy latitudes causes aerosol n.s.s. sulphate loadings to be obscured in bulk aerosol samples by high concentrations of sea-salt sulphate^{6,7}. To avoid this problem a two-stage sampler was used, comprising a 47-mm Nuclepore filter (8-μm pore size) upstream of a 47-mm Fluoropore filter (1-μm pore size). Operation at 30 l min⁻¹ gave a theoretical⁸ 50% cut-point for the Nuclepore filter at a particle radius of 1 μm, effectively excluding most of the sea salt from the Fluoropore filter, which thus collected the sub-micrometre sulphate aerosol component. (Selective sampling of the sub-micrometre particles is entirely appropriate because it is particles in this size range that dominate the CCN number concentration¹.) Similar sampling arrangements have been used successfully elsewhere²⁻⁴.

Aqueous extracts from the exposed Fluoropore filters were analysed for chloride, sulphate and MSA by ion chromatography (Dionex AS4A columns), to an estimated accuracy of ±5%. Filter blanks were always <5% of sample values for chloride and sulphate, and below detection limit for MSA; although we have made corrections for the blanks, the influence on the results was negligible. We estimated n.s.s. sulphate from chloride and sulphate concentrations assuming that the SO₄²⁻:Cl⁻ mole ratio in sea water is 0.0517 (ref. 9). We checked the validity of this by analysing approximately half of the filter extracts for sodium, also to ±5% precision, by atomic absorption spectroscopy. Figure 1 compares n.s.s. sulphate estimated from chloride to n.s.s. sulphate estimated from sodium, assuming a SO₄²⁻:Na⁺ molar ratio for sea water of 0.0603 (ref. 9). The results show that it is valid to use chloride as the seawater reference in our clean, marine samples.

Atmospheric DMS was determined by gas chromatography with flame photometric detection, using a wide-bore capillary

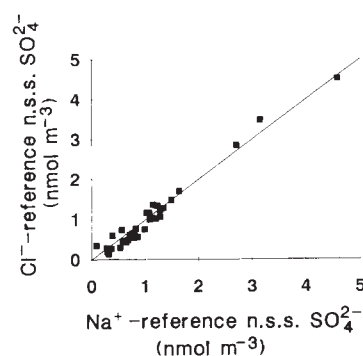


FIG. 1 Plot of n.s.s. SO₄²⁻ estimated using Cl⁻ as the seawater reference ion against n.s.s. SO₄²⁻ estimated using Na⁺ as the reference. Solid line has slope of one.

column (J & W scientific DB1 30 m $\times$ 0.53 mm $\times$ 5.0 μ m). DMS was preconcentrated on a gold-coated glass-wool plug, similar to the method first described by Ammons¹⁰ and later used by Andreae *et al.*¹¹, which used a gold-wool plug. Like Bates *et al.*² we used an alkali-impregnated (in our case NaOH) prefilter to remove oxidants before trapping the DMS. Calibrations using two different permeation tubes with independently determined loss rates agreed to within 10%.

The gold traps were cleaned at Australian Government Analytical Laboratories (AGAL) by heating them in a hydrogen stream at 450 °C. Clean traps were shipped to Cape Grim where the DMS samples were taken in baseline conditions for 1 h at a flow rate of 1 l min⁻¹. Exposed traps were returned to AGAL overnight, and the DMS determined within a day. Traps were exposed alternately at 2 a.m. and 1 p.m. (local time) to avoid any bias resulting from a diurnal DMS cycle when comparing the short-duration DMS data with the weekly average filter data. No diurnal cycle is discernible in the data collected so far.

Transportation problems resulted in DMS traps being exposed at irregular intervals, at considerably less than the twice-weekly frequency planned, and with gaps of weeks at a time during the summer of 1989–1990. Nevertheless, the data are sufficient to outline the DMS seasonal cycle, as shown in Fig. 2. The seasonal cycle is pronounced, with a mid-summer maximum and a mid-winter minimum, in keeping with expectations based on seawater and atmospheric DMS data from mid to high latitudes in the Northern Hemisphere^{5,12,13}, as well as seawater DMS data from the coast of Antarctica¹⁴. The latter showed a clear association between the mid-summer maximum in seawater DMS concentration and the concentrations of the alga *Phaeocystis pouchetii* and the algal metabolic product dimethylsulphoniopropionate, which is thought to be the precursor to DMS¹⁴. Our data suggest that a similar mid-summer algal bloom occurs in the waters off Cape Grim. The amplitude of the seasonal cycle at Cape Grim is a factor of about 10–20: DMS concentration ranged from 0.5–1 nmol m⁻³ in the winter of 1989 to 6–10 nmol m⁻³ in the two summer peaks (although there was a sampling gap of nearly a month at the expected time of the 1989–1990 maximum). Although our DMS data are based on only two summer peaks, we believe that they are representative

because the amplitude and phase of the cycle measured at Cape Grim is virtually identical to that measured¹⁵ in two other years at Amsterdam Island, 5,600 km due west of Cape Grim at 38° S in the southern Indian Ocean.

Figure 2c shows the aerosol MSA time series at Cape Grim. The amplitude, phase and shape of the MSA curve is sufficiently similar to that of DMS to provide compelling evidence that the two series are tightly coupled. The multiyear MSA data available from high-volume aerosol samples at Cape Grim⁶ exhibit a seasonal cycle identical to that in Fig. 2 (but with a concentration factor of about two higher because the high-volume sampler does not segregate the sample into super- and sub-micrometre size fractions). This confirms that our more limited time series is representative. The amplitude of the seasonal cycle for the sub-micrometre aerosol MSA data in Fig. 2 is similar to that for DMS, a factor of 12–25, going from $\sim$ 0.02 nmol m⁻³ mid-winter to 0.25–0.5 nmol m⁻³ in mid-summer.

The sub-micrometre n.s.s. sulphate data in Fig. 2b also show a strong seasonal cycle, with a mid-winter minimum and mid-summer maxima. The shape of the seasonal variation is not identical to that of the MSA curve, however, and there is more noise in the n.s.s. sulphate curve. The amplitude of the seasonal cycle for n.s.s. sulphate also seems to be somewhat smaller than that of the DMS and MSA curves, ranging over a factor of only about 5–10, from $\sim$ 0.3 nmol m⁻³ in mid-winter to 1.5–3 nmol m⁻³ in mid-summer (we acknowledge, however, that these estimates cannot be made precisely). On the other hand, much of the detailed structure (spikes) on the MSA curve is reproduced in the n.s.s. sulphate curve, especially in the two summer periods where DMS concentrations are high. Our conclusion is that aerosol n.s.s. sulphate concentrations, like those of MSA, are strongly coupled to DMS, especially at the summer maximum where the DMS concentration is about three times that of the combined concentrations of sub-micrometre MSA and n.s.s. sulphate.

But the difference in the shapes of the cycles in MSA and n.s.s. sulphate seems to be significant, as the time series of the ratio of MSA and n.s.s. sulphate in Fig. 2d shows. Despite the noise in the data, a distinct annual cycle in the ratio is evident, with apparently the same phase as the DMS, MSA and n.s.s. sulphate curves. The ratio ranges from a mean of $\sim$ 0.06 (range

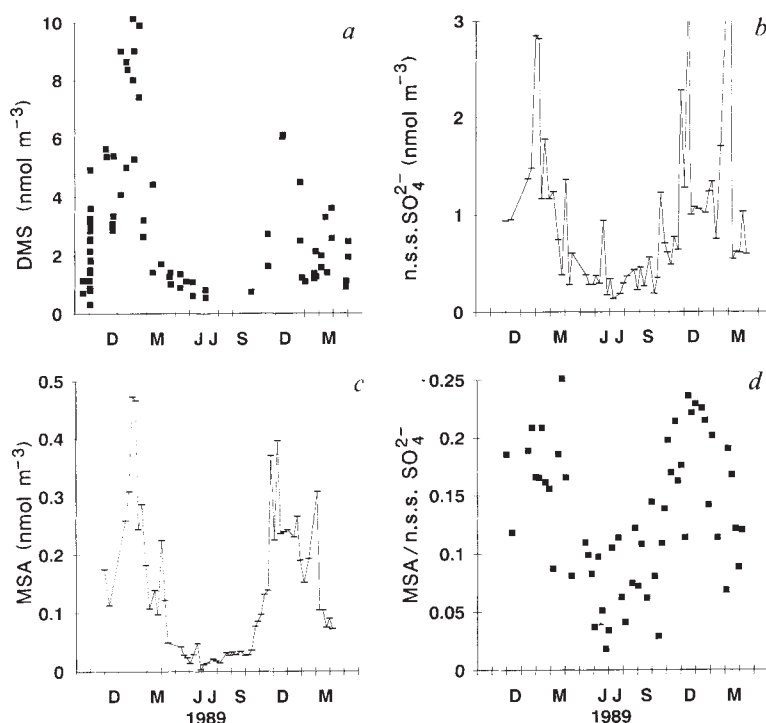


FIG. 2 Time-series plots of *a*, atmospheric DMS concentration; *b*, sub-micrometre-aerosol n.s.s. SO₄²⁻ concentration; *c*, sub-micrometre-aerosol MSA concentration; *d*, ratio of MSA and n.s.s. SO₄²⁻ in 'baseline' air at Cape Grim.

0.02–0.1) in mid-winter to ~ 0.18 (range 0.12–0.24) in mid-summer, for an average amplitude of a factor of about three.

The nonlinearity in the seasonal cycles of MSA and n.s.s. sulphate may imply the existence of another source of aerosol sulphur, in addition to DMS. If such a source did not have as distinct a winter minimum as DMS, it would be relatively more important in winter and could account for the reduced ratio of MSA to n.s.s. sulphate in winter. Without concurrent data on other sulphur species, we cannot pursue such a possibility here. Nevertheless, the decade of continuous data on condensation nuclei and CCN from Cape Grim^{16,17} show strong seasonal cycles in concentration, in sympathy with those of the three sulphur species discussed here, but with amplitudes less than that found here for DMS and MSA. For condensation nuclei¹⁶ the amplitude is a factor of about five, and for CCN at 1.2% supersaturation the amplitude is a factor of about four, decreasing to a factor of two at 0.23% supersaturation¹⁷. These amplitudes are more like that of our n.s.s. sulphate data, perhaps supporting the suggestion that there may be a significant non-DMS source of aerosol sulphur in winter, but also show nonlinearities as a function of supersaturation (a surrogate for particle size) that further complicate the picture.

The ratio of MSA to n.s.s. sulphate has been quoted⁷ as being rather constant at ~ 0.065 in high-volume aerosol samples from tropical and sub-tropical marine sites, but has consistently been reported to be between ~ 0.2 and >0.5 at southern mid to high latitudes in short-duration field campaigns^{2–4}, and ~ 0.2 in Antarctic ice¹⁸. Our data, revealing a summer maximum and winter minimum in the ratio of MSA to n.s.s. sulphate, are contrary to the explanation most commonly advanced to explain the differences between the tropical/sub-tropical data⁴ and the data from southern mid to high latitudes. That explanation hinges on a strong temperature dependence of the efficiency by which OH radical oxidizes DMS in the atmosphere to produce aerosol sulphur species. The overall oxidation pathways and intermediate species are not quantitatively understood¹⁹, but it is known that OH attacks DMS initially either by abstraction of a proton or by addition to the sulphur atom^{20,21}. The relative efficiencies of these pathways change with temperature, leading Berresheim^{3,4,22} to suggest that if MSA is a major product from the addition pathway, which available data shows to be favoured at low temperatures, then this could account for the higher ratio of MSA to n.s.s. sulphate in the data from the cooler oceans of the mid to high latitudes. This explanation is at odds with our data, because the summer-to-winter temperature variation at Cape Grim of 16 to 9 °C would change the efficiencies of the addition and abstraction pathways²⁰ from $0.27/0.73 = 0.37$ to $0.55/0.45 = 1.2$, which would predict that the ratio of MSA to n.s.s. sulphate would increase by a factor of three from summer to winter, whereas we find (Fig. 2d) that the ratio decreases by a factor of three.

Clearly, extended series of measurements involving a range of other DMS oxidation products, such as SO₂, are required if the marine sulphur cycle is to be quantitatively understood. One difficulty is that data obtained by different groups must be compared: in the past, some of the differences between data sets may have arisen from biases caused by differences in sampling technique. The tropical data which yielded low ratios of MSA to n.s.s. sulphate were taken with high-volume aerosol samplers, which collect a bulk aerosol sample on a single filter, whereas the high values reported from Southern Hemisphere studies have been produced by dual-filter samplers which segregate the aerosol into separate coarse and fine size fractions. The high loading of sea salt, which is alkaline, caused the high-volume filters used to determine the ratio of MSA to n.s.s. sulphate in tropical regions to pick up acidic gases such as HNO₃ (ref. 23)—it would be surprising if SO₂ were not similarly collected by the filters. Given that SO₂ is a known oxidation product of DMS, and exists in unpolluted marine air at concentration levels similar to those of n.s.s. sulphate^{2–4,24}, collec-

tion of SO₂ (which appears as sulphate in the analysis) along with n.s.s. sulphate on high-volume filters would lead to an apparent halving of the actual ratio of MSA to n.s.s. sulphate. In contrast, dual-filter arrangements such as used at Cape Grim should collect the SO₂ on the front filter, protecting the rear filter from artefact sulphate formation.

Pszenny *et al.*²⁴ have presented data showing the ratio of MSA to n.s.s. sulphate to be size-dependent, so bias may be introduced by the different aerosol cut-offs in different samplers. There may also be different seasonal variations in the size distribution of MSA and n.s.s. sulphate, complicating the comparison of results from different studies collected in different seasons. Concurrent measurements of size-distributed aerosol composition and gas-phase concentrations of at least DMS and SO₂, and preferably a number of other sulphur gases, over seasonal timescales are required. □

Received 3 August; accepted 21 November 1990.

1. Charlson, R. J., Lovelock, J. E., Andrea, M. O. & Warren, S. G. *Nature* **326**, 655–661 (1987).
2. Berresheim, H., Andrea, M. O., Ayers, G. P. & Gillett, R. W. in *Biogenic Sulfur in the Environment* (eds Cooper, W. J. & Saltzman, E. S.) 352–366 (American Chemical Society, Washington, DC, 1988).
3. Bates, T. S. *et al.* *J. Atmos. Chem.* **10**, 59–81 (1990).
4. Berresheim, H. *et al.* *J. Atmos. Chem.* **10**, 341–370 (1990).
5. Bates, T. S., Cline, J. D., Gammon, R. H. & Kelly-Hansen, S. R. *J. Geophys. Res.* **92**, 2930–2938 (1987).
6. Ayers, G. P., Ivey, J. P. & Goodman, H. S. *J. Atmos. Chem.* **4**, 173–185 (1986).
7. Savoie, D. L. & Prospero, J. M. *Nature* **339**, 685–687 (1989).
8. Manton, M. J. *Atmos. Envir.* **12**, 1669–1675 (1978).
9. Millero, F. J. *A. Rev. Earth planet. Sci.* **2**, 101–150 (1974).
10. Ammons, J. M. thesis, Univ. of South Florida (1980).
11. Andrea, M. O. *et al.* *J. Geophys. Res.* **90**, 12891–12900 (1985).
12. Turner, S. M., Malin, G., Liss, P. S., Harbort, D. S. & Holligan, P. M. *Limnol. Oceanogr.* **33**, 364–375 (1988).
13. Leck, C., Larsson, U., Bagander, L. E., Johansson, S. & Hajdu, S. *J. Geophys. Res.* **95**, 3353–3363 (1990).
14. Gibson, J. A. E., Garrick, R. C., Burton, H. R. & McTaggart, A. R. *Mar. Biol.* **104**, 339–346 (1990).
15. Nguyen, B. C., Mihalopoulos, N. & Belviso, S. *J. Atmos. Chem.* **11**, 123–141 (1990).
16. Gras, J. L. *J. Atmos. Chem.* **11**, 89–106 (1990).
17. Gras, J. L. *Geophys. Res. Lett.* **17**, 1565–1567 (1990).
18. Ivey, J. P., Davies, D. M., Morgan, V. & Ayers, G. P. *Tellus B* **38**, 375–379 (1986).
19. Plane, J. M. C. *ACS Symp. Ser.* **393**, (1988).
20. Hynes, A. J., Wine, P. H. & Semmes, D. H. *J. Phys. Chem.* **90**, 4148–4156 (1986).
21. Barnes, I., Bastian, V. & Becker, K. H. *Int. J. Chem. Kinet.* **20**, 415–431 (1988).
22. Berresheim, H. *J. Geophys. Res.* **92**, 13245–13262 (1987).
23. Prospero, J. M. & Savoie, D. L. *Nature* **339**, 687–689 (1989).
24. Pszenny, A. A. P., Castelle, A. J., Galloway, J. N. & Duce, R. A. *J. Geophys. Res.* **94**, 9818–9830 (1989).

ACKNOWLEDGEMENTS. We thank L. Porter and M. Leonard for supervision of the sampling, and P. Kitter and T. Lewis for the analysis. This work was supported by the Cape Grim research fund, administered jointly by the Australian Bureau of Meteorology and CSIRO.

Global nature of the Younger Dryas cooling event inferred from oxygen isotope data from Sulu Sea cores

H. R. Kudrass*, H. Erlenkeuser†, R. Vollbrecht‡ & W. Weiss§

*§ Bundesanstalt für Geowissenschaften, Stilleweg 2, 3000 Hannover 51, Germany

† Institut für Reine und Angewandte Kernphysik, Leibnizstrasse 19, 2300 Kiel, Germany

‡ Institut für Geologie und Paläontologie, Goldschmidt-Strasse 3, 3400 Göttingen, Germany

THE Younger Dryas, an approximately 1,000-year-long return to near-glacial conditions, interrupted the glacial/Holocene climate transition during which most of the Northern Hemisphere ice sheets melted. Evidence for the Younger Dryas event has been found mainly in sediments from the North Atlantic Ocean and northwest Europe, and this has led to the idea that the event was caused by an injection of meltwater into the North Atlantic Ocean^{1–3}. This model, however, has been recently questioned in the light of coral-reef data on the rate of sea level changes during this transition⁴. Here we present high-resolution oxygen isotope

records from benthic and planktonic foraminifera from two radiocarbon-dated cores in the Sulu Sea, showing that the Younger Dryas occurred synchronously in the surface and deep waters of the Sulu Sea and the northern Atlantic Ocean. By combining our results with other palaeoclimate data, we suggest that the Younger Dryas event was a global phenomenon, and we believe it to have been caused by low atmospheric CO₂ concentrations.

The Younger Dryas has attracted special attention^{1,2,4-10}. Radiocarbon dating using accelerator mass spectrometry (AMS) of foraminiferal tests^{1,11} have fixed this event between ~11,000 and 10,000 radiocarbon yr BP. The Younger Dryas is usually referred to as a regional climate and oceanographic phenomenon, confined to the northern Atlantic Ocean and the surrounding land areas¹⁻³, and the Gulf of Mexico¹².

The Sulu Sea is a marginal oceanic basin with maximum water depths of ~5,000 m in the Sulu Trench. Water exchanges with the neighbouring South China and Celebes seas through shallow straits, the deepest of which is Mindoro Strait (sill depth 420 m). The surface water, which is driven by monsoonal winds, mixes easily with that of the surrounding seas¹³. In contrast, the Sulu Sea deep water is renewed by episodic surges of South China Sea Intermediate Water (600–1,000 m) through the Mindoro Strait¹⁴. As a result, temperatures (9.8–10.0 °C) and salinities (34.48‰) are remarkably constant below 1,000 m^{13,14}.

Sedimentation rates average 5.5 cm kyr⁻¹ for purely hemipelagic sediments^{15,16}, but may exceed 60 cm kyr⁻¹ for turbidites in the Sulu Trench¹⁷. We chose two cores from the western margin of the Sulu Trench, which consist of millimetre-thin layers of hemipelagic sediment separated by about 200, centimetre-thin, distal, laminated mud turbidites deposited at an average frequency of one every 70 years¹⁴. (Core locations—SO49-KL82: 121°38.19' E, 8°9.36' N, 4,911 m water depth; SO58-KL69: 121°35.7' E, 8°49.5' N, 4,696 m water depth.) These cores allow us to analyse the isotopic changes of the Sulu Sea water since the Last Glacial Maximum (~18 kyr ago), because every layer of hemipelagic sediment sandwiched between two turbidites has been largely protected from bioturbation. The isotopic compositions of foraminiferal tests from layers of hemipelagic sediment therefore represent the short time intervals between turbidite events. Hence the isotopic changes in surface water are well preserved.

The stable isotopic composition was determined by the auto-

matic CARBO Kiel/MAT 251 CO₂-preparation and mass spectrometer system. Between 10 and 20 tests of the planktonic foraminifer *Globigerinoides ruber* white (>215 µm) and 2–5 tests of the benthic foraminiferal genus *Bulimina* sp., recovered from layers of hemipelagic sediment, were ultrasonically cleaned and reacted with 100% H₃PO₄ under vacuum. Isotope ratios are given relative to the PDB Standard, with a total accuracy of 0.05‰ and 0.08‰ for δ¹³C and δ¹⁸O, respectively. For conventional radiocarbon dating, mixed species of planktonic foraminifera (>125 µm) from one, or sometimes two, neighbouring hemipelagic layers were separated and analysed in ¹⁴C miniature counters¹⁸.

The planktonic isotope records of the two cores studied agree remarkably well if the generally higher sedimentation rate in core SO58-KL69 is taken into account (Fig. 1). The δ¹⁸O values decrease from maximum glacial values of -1.1‰ to minimum Holocene values of -2.7‰. This shift from glacial to interglacial conditions is clearly interrupted by a return to near-glacial values that we take to be the equivalent of the Younger Dryas cooling event. The δ¹⁸O increase during this interval is more prominent in core SO49-KL82 than in core SO58-KL69 (peak maximum of -1.2 and -1.6‰, respectively). This probably results from the higher frequency of turbidites in core SO49-KL82, which improves the resolution of the signal by reducing bioturbational mixing.

Considering a pure ice-volume effect of 0.6‰ for the time of the Younger Dryas⁴, and subtracting that effect from the maximum Younger Dryas signal in the high-resolution core SO49-KL82 (Fig. 1c), a residual value of -1.8‰ is obtained. The difference of 0.9‰ from present isotope values of -2.7‰ indicates changes in the surface water of the Sulu Sea that correspond to a drop in temperature of ~3 °C or, alternatively, a reduction of salinity by 1‰, or a combination of both.

The species composition of planktonic foraminifera also reflects environmental changes of surface waters. During the Younger Dryas, *Globigerinoides ruber* was as abundant (>20%) as during the Last Glacial Maximum, but occurred less frequently (<12%) during the periods preceding and succeeding the Younger Dryas. This suggests that the Younger Dryas climate conditions were comparable to those of the Last Glacial Maximum, which in the adjacent South China Sea, were characterized by 5 °C lower temperatures than today (based on analyses

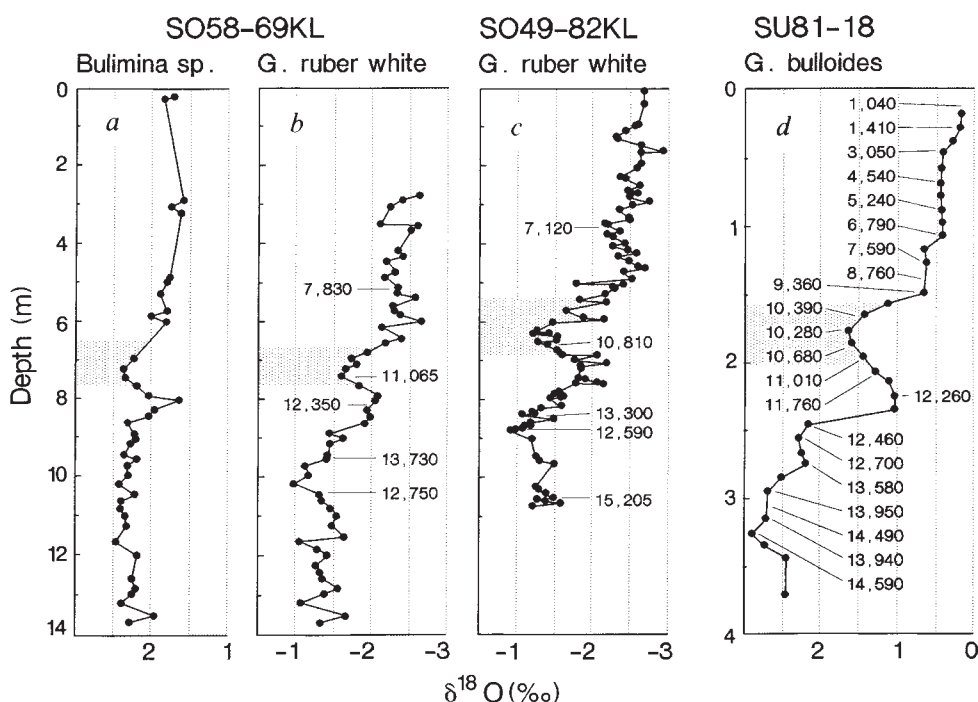


FIG. 1 Radiocarbon-dated oxygen isotope record of two cores from the Sulu Sea (SO58-69KL: 13.72 m long, SO49-82KL: 10.88 m) and one core from the North Atlantic Ocean (SU81-18: 3.80 m; ref. 11). The Younger Dryas interval is indicated by the shaded area. ¹⁴C dates of planktonic foraminifera have been corrected for the surface-water age of 400 yr. a, Record for benthic foraminifera *Bulimina* sp.; b and c, record for planktonic foraminifera *Globigerinoides ruber*; d, record for planktonic foraminifera *Globigerina bulloides*¹¹.

TABLE 1 ^{14}C dates of planktonic foraminifera ($>125\ \mu\text{m}$)

Number of core	Depth in core (cm)	Weight of sample (g)	^{14}C age (yr)
SO49-KL82	346.0–350.5	0.34	$7,120 \pm 245$
SO49-KL82	653.0–653.5	0.80	$10,810 \pm 305$
SO49-KL82	833.5–837.0	1.90	$13,300 \pm 390$
SO49-KL82	873.5–878.0	0.88	$12,590 \pm 350$
SO49-KL82	1,039.0–1,041.0	0.34	$15,205 \pm 525$
SO58-KL69	509.5–521.0	0.89	$7,830 \pm 245$
SO58-KL69	730.5–749.5	1.60	$11,065 \pm 325$
SO58-KL69	796.0–800.5	1.59	$12,350 \pm 345$
SO58-KL69	938.0–942.0	0.81	$13,730 \pm 460$
SO58-KL69	1,041.0–1,043.5	0.90	$12,750 \pm 300$

Conventional radiocarbon ages have been corrected for reservoir age of 400 yr (ref. 11). ^{14}C activities were determined using 15-cm³ miniature counters of the radiocarbon laboratory of the Niedersächsisches Landesamt für Bodenforschung, Hannover.

of planktonic foraminifera¹⁹). More direct evidence for a considerable cooling during the Younger Dryas comes from the increased abundance of the cool-water species *Neogloboquadrina pachyderma* in sediments of the Sulu Sea²⁰. In addition, the Younger Dryas southward shift of the polar front in the northwestern Pacific Ocean²¹ probably also affected the marginal seas further to the south. The observed $\delta^{18}\text{O}$ shift in our Younger Dryas core sections is therefore interpreted to be mainly caused by a decrease in sea surface temperatures.

As the deep water of the Sulu Sea is replenished by water from the South China Sea^{13,14}, the ^{18}O values of the benthic foraminifera record the isotopic changes of the Intermediate South China Sea Water which itself originates from Western North Pacific Intermediate Water¹³. The late Holocene oxygen isotope values of the benthic foraminifera genus *Bulimina* differ by 0.9‰ from those of the last glacial stage, corresponding fairly well to the pure ice-volume effect of 1.2‰ (ref. 4) and probably indicating that glacial deep-water temperature and salinity were not significantly different from present-day conditions.

During the glacial/interglacial transition, benthic and planktonic isotope records generally vary in phase (Fig. 1). The benthic signal, however, shows a significant time lag at the onset of the first meltwater spike⁴ as indicated by the sharply decreasing isotope values between core depths of 8 m and 10 m (SO58-69 KL).

As the deep water of the Sulu Sea is replenished by intermediate water of the South China Sea¹⁴, which originates by sinking of cooled and dense surface waters during the winter season, the observed time lag in the benthic record may be explained by a reduced vertical transfer of the low-salinity, isotopically light meltwater-spike surface water. At the beginning of the Younger Dryas, the supply of meltwater was reduced⁴ and consequently both records change in phase, indicating a faster vertical exchange.

The isotope records were dated by conventional ^{14}C analyses (Table 1, Fig. 1). The variations correlate with AMS-radiocarbon-dated isotope records off southern Portugal¹¹ and off northern Japan²¹, where the Younger Dryas commenced at ~11,000 radiocarbon yr BP and ended ~10,000 radiocarbon yr BP. In the cores of the Sulu Sea, the lowermost section of the Younger Dryas, identified by the return to near-glacial isotope values, is dated as 10,800 and 11,065 radiocarbon yr BP. These dates correspond to the beginning of the Younger Dryas off Japan and Portugal.

The oxygen isotope variations, in combination with the ^{14}C dates, clearly show that the cooling event interrupting the glacial/interglacial transition in the Sulu Sea occurred at the same time as the Younger Dryas in the North Atlantic Ocean and northwest Europe. Considering further the ^{14}C -dated evidence for the Younger Dryas in the Gulf of Mexico¹², the North Pacific Ocean²¹ and Argentina²², as well as the shifts

presumably related to the Younger Dryas in oxygen isotope records from the equatorial Atlantic Ocean^{23,24} and the Bengal Fan²⁵, including also the Younger Dryas equivalent temperature changes recorded in Greenland²⁶ and Antarctic ice cores²⁷, the Younger Dryas event must be regarded as a global phenomenon.

Various climate models have been proposed to explain the origin of the Younger Dryas^{1–3,5,7,28}. Evidence is growing that the rapid climate changes during this brief (~1,000 yr) period were caused by feedback processes between the production of meltwater, the changing patterns of ocean circulation, and the CO_2 concentration of the atmosphere. During the warm period of the Allerød, preceding the Younger Dryas, the glacial climatic system became destabilized by increasing insolation, which caused large portions of the northern continental ice sheets to melt⁴. In high northern latitudes the meltwater produced a low-salinity surface layer and prevented the formation of cold North Atlantic Deep Water^{9,29,30}. This, in turn, may have kept the atmospheric CO_2 concentration at low glacial levels during the first meltwater pulse, that is, during the Allerød. Possibly, the inhibited exchange between atmosphere and deep oceanic waters, in combination with an increased consumption of CO_2 by plants (re-colonized ice-free areas, extended temperature zones) may even have led to a further decrease. The postulated continuation of low atmospheric CO_2 concentrations from the glacial maximum into the Younger Dryas does not conform to the increase of CO_2 concentrations in ice cores assigned to the end of the Last Glacial Maximum³¹. But with regard to the large uncertainties about the age of the entrapped CO_2 in relation to the age of the host ice, a later rise of CO_2 concentrations cannot be excluded. According to our conceptual model, the meltwater supply to the northern high-latitude ocean was reduced at that time because the reservoir of glacial ice which could be melted under climate conditions controlled by low atmospheric CO_2 concentration was exhausted. With a reduced meltwater supply at the beginning of the Younger Dryas, oceanic surface-mixing and deep-water circulation intensified^{9,29}, which, in turn, increased ocean productivity. These processes further decreased the low atmospheric concentration of CO_2 and, in contrast to a greenhouse effect, established the cold conditions of the Younger Dryas. This period was ended by the gradual release of CO_2 into the atmosphere from circulating deep water, which was controlled by the oceanic residence time of some hundred years. The prime factor for the climatic changes in this model for the Younger Dryas is the changing atmospheric CO_2 concentration. Analyses of gas compositions in ice cores combined with improved age determinations of entrapped CO_2 could probably specify the role of CO_2 as the cause of short-term climatic changes, possibly also of modern ones. □

Received 17 September; accepted 20 December 1990.

- Broecker, W. S. *et al.* *Paleoceanography* **3**, 1–19 (1988).
- Broecker, W. S. *et al.* *Nature* **333**, 156–158 (1988).
- Overpeck, J. T. *et al.* *Nature* **338**, 553–557 (1989).
- Fairbanks, R. G. *Nature* **342**, 637–642 (1989).
- Duplessy, J.-C. *et al.* *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **35**, 121–144 (1981).
- Ruddiman, W. F. & Duplessy, J.-C. *Quat. Res.* **23**, 1–17 (1985).
- Berger, W. H. *Episodes* **8**, 163–168 (1985).
- Dansgaard, W. *et al.* *Nature* **339**, 532–534 (1989).
- Jansen, E. & Veum, T. *Nature* **343**, 612–616 (1990).
- Sarnthein, M. & Tiedemann, R. *Paleoceanography* **5**, 1041–1055 (1990).
- Bard, E. *et al.* *Nature* **328**, 791–794 (1987).
- Kennett, J. P. *et al.* *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **50**, 189–216 (1985).
- Wyrtki, K. *NAGA-Report No. 2* (Univ. California, La Jolla, 1961).
- Quadtassel, D., Kudrass, H. R. & Frische, A. *Nature* **348**, 320–322 (1990).
- Linsley, B. K. *et al.* *Mar. Micropaleont.* **9**, 395–418 (1985).
- Vollbrecht, R. & Kudrass, H. R. in *Proc. ODP, Init. Rep. 124* (eds Rangin, C. *et al.*) 105–111 (Ocean Drilling Program, College Station, 1990).
- Exon, N. F. *et al.* *Mar. Geol.* **39**, 165–195 (1981).
- Jelen, K. & Geyh, M. A. *Radiocarbon* **28**, 578–585 (1986).
- Wang, L. & Wang, P. *Paleoceanography* **5**, 77–90 (1990).
- Linsley, B. K. & Thunell, R. C. *Paleoceanography* **5**, 1025–1039 (1990).
- Kallel, N. *et al.* *Oceanol. Acta* **11**, 369–375 (1988).
- Heusser, C. J. & Rabassa, J. *Nature* **328**, 609–611 (1989).
- Pastouret, L. *Oceanol. Acta* **1**, 217–232 (1985).
- Sarnthein, M., Erlenkeuser, H. & Zahn, R. *Bull. Inst. Geol. Bassin d'Aquitaine* **31**, 393–407 (1982).
- Duplessy, J.-C. *et al.* *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **33**, 9–46 (1981).
- Dansgaard, W. *et al.* *Science* **218**, 1273–1277 (1982).

27. Jozel, J. et al. *Nature* **329**, 403–408 (1987).
 28. Emiliani, C., Rooth, C. & Stipp, J. J. *Earth planet. Sci. Lett.* **41**, 159–162 (1978).
 29. Jansen, E. & Erlenkeuser, H. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **49**, 189–206 (1985).
 30. Dean, W. E. et al. *Paleoceanography* **4**, 639–653 (1989).
 31. Barnola, J. M. et al. *Nature* **329**, 408–414 (1987).

ACKNOWLEDGEMENTS. We thank M. Sarnthein for a careful review of an earlier version of the manuscript and M. Geyh for ^{14}C data. Analyses of stable isotopes were supported by the German National Climate Program.

A new high-pressure form of MgAl_2O_4

T. Irifune*, K. Fujino* & E. Ohtani†

* Department of Earth Sciences, Faculty of Science, Ehime University, Matsuyama 790, Japan

† Institute of Mineralogy, Petrology and Economic Geology, Faculty of Science, Tohoku University, Sendai 980, Japan

MAGNESIUM aluminium spinel (MgAl_2O_4) is a common constituent of low-pressure peridotite xenoliths, and is an important host mineral for aluminium and other trivalent cations in the shallow upper mantle. Shock-wave compression data for MgAl_2O_4 demonstrated¹ that spinel would transform to denser phases at pressures greater than 40 GPa, although there was much uncertainty in the pressure estimation. Later, spinel was found to transform under static compression to the denser oxide mixture (MgO periclase + Al_2O_3 corundum^{2,3}) at pressures above 15 GPa (ref. 4). Detailed analyses of the shock-compression data, however, suggested the presence of a still denser phase of MgAl_2O_4 spinel at much higher pressures^{5,6}. Here we report the transformation of MgAl_2O_4 spinel to a new high-pressure form at pressures above 25 GPa in a multi-anvil high-pressure apparatus. The new MgAl_2O_4 phase has a structure similar to that of CaFe_2O_4 (calcium ferrite) and its zero-pressure density is $3.937(3) \text{ g cm}^{-3}$, which is $\sim 2\%$ denser than the lower-pressure assemblage of periclase + corundum. We suggest that this high-pressure form of MgAl_2O_4 may be an important host of aluminium in the Earth's lower mantle.

Our high-pressure experiments on MgAl_2O_4 samples were performed using the 1,500-ton multi-anvil apparatus at Ehime University. Pressure was calibrated by measurements of the resistance changes associated with the room-temperature phase transitions of ZnS (15.6 GPa), GaAs (18.7 GPa) and GaP (23 GPa). Corrections for the effect of high temperature were also made on the basis of phase transitions in Mg_2SiO_4 (olivine to modified spinel) and MgSiO_3 (ilmenite to perovskite). Uncertainties in pressure measurements were around ± 0.5 GPa. Our starting material was single-crystal spinel of pure MgAl_2O_4 (supplied by Union Carbide Co., USA), which was crushed into a fine-grained powder (grain size of several micrometres), and enclosed in a platinum capsule. Temperature was measured with a Re3%W97%–Re25%W75% thermocouple, the hot junction of which was placed near the sample capsule. As the thermocouple junction is not directly in contact with the sample, the nominal temperatures may suffer uncertainties of about ± 50 – 100°C . Runs were quenched after 5–15 min heating under pressure and the products were examined by X-ray powder diffraction, scanning electron microscopy (SEM), electron probe microanalysis (EPMA) and transmission electron microscopy (TEM). Details of the high-pressure experiments have been reported elsewhere⁷.

Runs were carried out at pressures of 14, 15, 16 and 26 GPa at a temperature of $1,500^\circ\text{C}$. Additional runs were made at 16 and 25 GPa, $1,000^\circ\text{C}$. Spinel was found to decompose into a mixture of periclase and corundum at pressures above 15 GPa at both of these temperatures, in agreement with earlier results^{2–4}. X-ray analysis of the product at 25 GPa and $1,000^\circ\text{C}$, however, revealed the presence of several new lines in the powder diffraction pattern, in addition to those of periclase and corundum. The sample was found to convert mostly to this new phase at 26 GPa and $1,500^\circ\text{C}$, with a few diffuse lines of corun-

dum also evident. Periclase was difficult to identify because it gives very few diffraction lines and because the most intense line overlaps with a line for the new phase. No diffraction lines for the starting material (spinel) were found in these two runs. SEM observations for the run product of 26 GPa revealed that the sample was composed of crystalline aggregates, mostly 2–5 μm in size. The chemical composition of the new phase was almost exactly MgAl_2O_4 on the basis of EPMA analyses of relatively large grains (molar ratio of $\text{MgO}:\text{Al}_2\text{O}_3 = 1.02:0.98$). The small deviation from the ideal 1:1 (MgAl_2O_4) composition may be due to overlapping of the electron beam onto traces of adjacent crystals, because the crystals available were only just larger than the minimum required for chemical analysis by wavelength-dispersive EPMA.

X-ray powder diffraction data for the new MgAl_2O_4 phase are listed in Table 1. These data were taken with a 57.3-mm Debye–Scherrer camera using $\text{Cu K}\alpha$ radiation. Over 40 lines ($0^\circ < 2\theta < 90^\circ$) corresponding to the new phase were observed, a few of which overlap with those for corundum. The diffraction patterns for the new MgAl_2O_4 phase could be indexed on the basis of an orthorhombic cell with lattice parameters $a = 8.631(3) \text{ \AA}$, $b = 9.969(3) \text{ \AA}$, $c = 2.789(1) \text{ \AA}$ and volume $V = 240.00(15) \text{ \AA}^3$. The diffraction pattern is similar to that of CaFe_2O_4 calcium ferrite, for which intensity data are listed in Table 1. The characteristics of the X-ray reflection extinction are also consistent with the space group *Pham*. The zero-pressure density for this calcium ferrite-type MgAl_2O_4 phase was calculated as $3.937(3) \text{ g cm}^{-3}$ on the basis of four formula units in a unit cell ($Z = 4$), which is substantially larger than the density of spinel ($\rho_0 = 3.578 \text{ g cm}^{-3}$).

TEM observations have also been carried out for the sample produced at 26 GPa and $1,500^\circ\text{C}$. Electron diffraction spots were clearly observed for each grain of the new MgAl_2O_4 phase, and are consistent with the X-ray powder diffraction data. An example of these TEM observations is shown in Fig. 1.

These data suggest that MgAl_2O_4 spinel transforms to a new calcium ferrite-type structure at pressures above 25 GPa via a lower-pressure assemblage of periclase + corundum. Definitive identification of the structure of the new phase will require refinement of single-crystal data. It is difficult, however, to produce single crystals large enough for X-ray structural refinements at such high pressure. We are currently pursuing this.

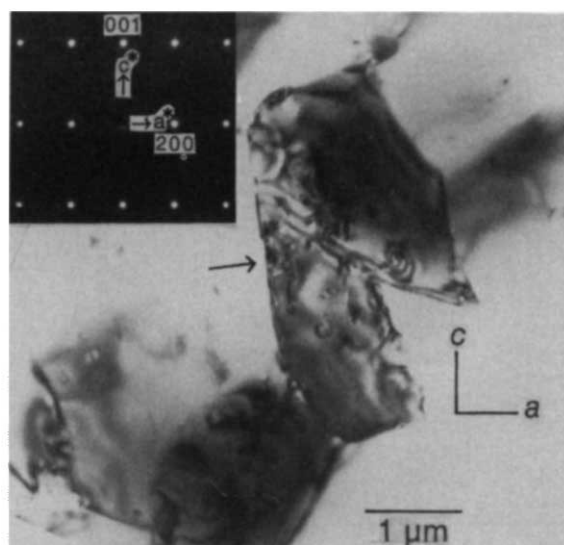


FIG. 1 Transmission electron micrograph of the new calcium ferrite-type MgAl_2O_4 produced at 26 GPa and $1,500^\circ\text{C}$. Scale bar, 1 μm . Inset is the selected-area electron diffraction pattern of the grain indicated by the arrow. The diffraction spots are reasonably indexed on the basis of the lattice parameters determined by X-ray powder diffraction.

It has been suggested^{5,6} that spinel would transform to an unknown phase denser than the oxide mixture ($\rho_0 = 3.864$). The calcium ferrite structure had been considered a candidate for this phase^{6,8}, until proved unlikely by the discovery of an extremely dense phase using a diamond-anvil cell⁹. This phase, tentatively labelled ϵ -MgAl₂O₄, was reported to have an orthorhombic cell with a zero-pressure density of 4.310 g cm⁻³ and to have been formed at pressures greater than 25 GPa, immediately above the stability field of the oxide mixture. Although the pressure and temperature conditions of that study are similar to those used here, our results show no evidence of this ϵ -MgAl₂O₄ phase in the run products.

We now discuss possible reasons for this inconsistency between the two sets of results. ϵ -MgAl₂O₄ was synthesized in a diamond-anvil cell at pressures of 25 and 28 GPa, which was calibrated on the basis of the NaCl scale at room temperature⁹. It is supposed, however, that these pressures are probably under-

estimated by about 20% in high-temperature runs⁹. This, and the fact that the ϵ -MgAl₂O₄ phase possesses a higher zero-pressure density than the new calcium ferrite phase, suggest that ϵ -MgAl₂O₄ may be a high-pressure form of the latter phase.

Alternatively, it is possible that one of these phases is metastable. The ϵ -MgAl₂O₄ phase was synthesized from a reactive gel of composition MgAl₂O₄, whereas we used crystalline spinel as the starting material. Using a gel as the starting material has the merit of promoting fast reaction times. This, however, may also cause rapid growth of metastable phases as temperature is increased, because of the lower temperatures and probably higher deviative stresses relative to those at the final stage of the heating experiments under compression. Once a metastable phase is formed, it can be difficult for it to transform to the more stable phase in the short reaction time involved in ordinary laser-heating experiments. Thus it is preferable to use crystalline starting materials as long as reactions proceed efficiently during the heating time available. In the present case, spinel was found to transform completely to the high-pressure phases at pressures above 25 GPa at both 1,000 and 1,500 °C, which suggests that reactions do indeed proceed rapidly under such pressure-temperature conditions.

Shock-wave compression experiments have been performed recently¹⁰ for synthetic spinel crystals at pressures of 25.5–50.5 GPa. TEM observations of the product indicated that spinel partially transformed to an unidentified phase which is probably not the ϵ -MgAl₂O₄ phase. Detailed characteristics of this phase have not been reported yet, so it is unclear whether or not it corresponds to the calcium ferrite-type MgAl₂O₄. Further experimental studies are clearly necessary to understand the stability of the new MgAl₂O₄ phase and the relation between this and the ϵ -MgAl₂O₄ phase at higher pressures.

The ϵ -MgAl₂O₄ phase gives a better fit⁹ to the shock Hugoniot data than the assumed calcium ferrite phase⁶ ($\rho_0 = 4.13$ g cm⁻³). This analysis, however, yielded unreasonably high bulk moduli, K_0 (540 GPa and 400 GPa respectively) and low pressure derivatives of the bulk moduli, K'_0 (0.6 and 1.5) for these phases relative to those observed for most minerals in the mantle, including their high-pressure phases (normally $K_0 = 200$ –300 GPa and $K'_0 \approx 4$). The lower zero-pressure density that we observe for the calcium ferrite phase would result in more reasonable values of K_0 and K'_0 . But such analyses carry considerable uncertainties, and need to be determined from measurements of the elastic properties in static-compression experiments.

In the calcium ferrite structure, the most dense structure for known crystalline materials with AB₂O₄ stoichiometry⁸, Mg²⁺ ions should be in 8-coordinated sites and Al³⁺ ions occupy octahedral sites. NaAlSiO₄ also adopts this structure at 1,000 °C and pressures above 18 GPa (ref. 11). On the other hand, a recent high-pressure study⁷ demonstrated that Al³⁺ and Na⁺ ions are incorporated into majorite garnet at pressures corresponding to the transition region (400–670 km depth) of the mantle, so the calcium ferrite structure may not play an important role in this region. Under lower-mantle conditions, majorite garnet is no longer stable^{7,12}, and other dense phases are expected to occur there instead. Several possible candidates have been proposed for the aluminium-bearing phase in the lower mantle: ϵ -MgAl₂O₄ (ref. 9), aluminous MgSiO₃ and CaSiO₃ perovskites¹³, an unidentified 'Al-rich phase'¹², and a hollandite-type structure¹⁴ (Ca_{0.5}Mg_{0.5})Al₂Si₂O₈. Our study suggests that the calcium ferrite-type MgAl₂O₄ is another potential host for Al³⁺ and probably for Na⁺. In the absence of experimental data on the detailed phase equilibria and mineralogical characteristics of these phases, it is not possible at present to decide which phase is actually dominant in the lower mantle. To solve this problem, static high-pressure and high-temperature data are needed for more realistic mantle compositions and at pressures greater than 25 GPa, coupled with the precise characterization of the product phases.

TABLE 1 X-ray powder diffraction data for the new MgAl₂O₄ phase

d_{obs}	d_{cal}	l/l_0	$(l/l_0)^*$	$h\ k\ l$
5.00	4.98	<5	(2)	0 2 0
4.310	4.316	30	(20)	2 0 0
3.260	3.263	15	(9)	2 2 0
3.091	3.101	5	(2)	1 3 0
2.761	2.764	10	(4)	3 1 0
2.694	2.686	<5	(6)	0 1 1
2.635	2.633	5	(2)	2 3 0
C	2.565	—	(4)	1 1 1
2.491	2.492	100	(100)	0 4 0, 3 2 0
C	2.394	—	(4)	1 4 0
2.345	2.343	30	(70)	1 2 1, 2 0 1
2.176	2.175	20	(2)	3 3 0
2.159	2.158	5	(2)	2 4 0, 4 0 0
2.112	2.109	10	(4)	4 1 0
2.074	2.074	20	(24)	1 3 1
1.981	1.980	30	(16)	4 2 0
1.963	1.963	10	(20)	3 1 1
1.941	1.943	10	(3)	1 5 0
1.919	1.915	<5	(2)	2 3 1
1.860	1.858	5	(2)	3 2 1
1.812†	1.817	20	(9)	1 4 1
	1.810		(1)	2 5 0
1.711	1.715	5	(10)	3 3 1
1.7063	1.7068	40	(77)	2 4 1, 4 0 1
1.6817	1.6822	15	(10)	4 1 1
1.635†	1.6388	10	(2)	3 5 0
	1.6316		(6)	1 6 0, 4 4 0
C	1.6147	—	(2)	4 2 1
1.5503	1.5506	30	(10)	2 6 0
1.5176	1.5182	10	(4)	2 5 1, 4 3 1
1.4386	1.4387	30	(25)	3 6 0, 6 0 0
1.422†	1.4238	5	(2)	6 1 0
	1.4191		(6)	5 4 0
1.4079	1.4083	35	(27)	1 6 1
1.3934	1.3946	5	(19)	0 0 2
1.3537	1.3553	10	(5)	2 6 1
1.3181	1.3165	10	(1)	4 6 0
1.2772	1.2786	5	(2)	3 6 1, 6 0 1
1.2469	1.2460	15	(2)	0 8 0, 6 4 0
1.2336	1.2334	10	(3)	1 8 0
1.2170	1.2170	10	(14)	3 2 2
1.1984	1.1971	5	(3)	7 2 0, 2 8 0
1.1908	1.1905	5	(2)	4 6 1
1.1644	1.1666	<5	(1)	6 5 0
1.1411	1.1402	<5	(1)	4 2 2

Relative intensities for the new MgAl₂O₄ phase were estimated visually. Cell parameters: $a = 8.631(3)$, $b = 9.969(3)$, $c = 2.789(1)$ Å, $V = 240.00(15)$ Å³.

* Intensities for CaFe₂O₄ calcium ferrite taken from the JCPDS card (No. 32-168).

† Broad lines; C, lines overlapped with the diffraction lines for corundum, and were therefore not used for lattice-constant refinement.

Provided that the lower mantle is of pyrolitic composition, mass-balance considerations suggest that if the new MgAl_2O_4 phase does exist in the mantle, it comprises probably less than 10% of the lower-mantle material. Its presence would therefore cause no marked changes in lower-mantle seismic velocities. Nevertheless, the slight increase in seismic velocities observed at depths of 670–800 km (ref. 15) may be relevant to such a transformation. Moreover, the formation of even a small amount of the new aluminous phase may influence significantly the behaviour of some trace elements such as rare-earth elements, with important consequences for differentiation processes in the early Earth¹⁶. MgAl_2O_4 spinel is one of the early condensed minerals and is commonly found in stony meteorites. Thus we suggest that shock-metamorphosed meteorites may contain this new high-pressure form of spinel, as was the case for olivine (ringwoodite) and pyroxene (majorite).

Note added in proof. It has been reported very recently¹⁷, on the basis of *in situ* X-ray measurements in a diamond-anvil cell¹⁷, that under lower-mantle conditions the major constituents of a natural peridotite are two perovskites and magnesio-wüstite. No chemical analyses have been performed on these run products, however, and we suppose this method may overlook the presence of relatively small amounts of aluminous and other phases. □

1. Birch, F. in *Handbook of Physical Constants* (ed. Clark, S. P. Jr) 97–173 (Geological Society of America, New York, 1966).
2. Liu, L.-G. *Geophys. Res. Lett.* **2**, 9–11 (1975).
3. Ohtani, E. et al. in *Proc. 4th Int. Conf. on High Pressure* (ed. Osugi, J.) 185–189 (Physical Chemistry Society of Japan, Tokyo, 1975).
4. Liu, L.-G. *High Temp.-High Press.* **12**, 217–220 (1980).
5. Wang, C. *Nature* **218**, 560–561 (1968).
6. Davies, G. F. & Gaffney, E. S. *Geophys. J. R. astr. Soc.* **33**, 165–183 (1973).
7. Irifune, T. & Ringwood, A. E. in *High-Pressure Research in Mineral Physics* (ed. Manghnani, M. H. & Syono, Y.) 231–242 (Terra Publ./American Geophysical Union, Tokyo, 1987).
8. Reid, A. F. & Ringwood, A. E. *Earth planet. Sci. Lett.* **6**, 205–208 (1969).
9. Liu, L.-G. *Earth planet. Sci. Lett.* **41**, 398–404 (1978).
10. Schäfer, H. et al. *Phys. Chem. Miner.* **9**, 248–252 (1983).
11. Liu, L.-G. *Geophys. Res. Lett.* **4**, 183–186 (1977).
12. Ito, E. & Takahashi, E. in *High-Pressure Research in Mineral Physics* (ed. Manghnani, M. H. & Syono, Y.) 221–230 (Terra Publ./American Geophysical Union, Tokyo, 1987).
13. Irifune, T. & Ringwood, A. E. *Earth planet. Sci. Lett.* **86**, 365–376 (1987).
14. Madon, M. et al. *Nature* **342**, 422–425 (1989).
15. Dziewonski, A. M. & Anderson, D. L. *Phys. Earth planet. Inter.* **25**, 297–356 (1981).
16. Kato, T. et al. *Earth planet. Sci. Lett.* **89**, 123–145 (1988).
17. O'Neil, B. & Jeanloz, R. *Geophys. Res. Lett.* **17**, 1477–1480 (1990).

ACKNOWLEDGEMENTS. We thank the Ministry of Education, Science, and Culture of the Japanese Government for financial support by Grant-in-Aid for Scientific Research. We also thank A. Yoneda, H. Ohashi, H. Sawamoto and A. E. Ringwood for their help and comments. Thanks are due also to T. S. Duffy and C. B. Agee for reviews.

Vertical disparities and perception of three-dimensional shape

B. G. Cumming, E. B. Johnston & A. J. Parker

University Laboratory of Physiology, Parks Road, Oxford OX1 3PT, UK

THE information about depth and three-dimensional shape available from the horizontal component of the stereo disparity field requires interpretation in conjunction with information about ego-centric viewing distance (D). A novel computational approach for estimating D was proposed by Mayhew and Longuet-Higgins^{1,2}, who demonstrated that the horizontal gradient of vertical disparities uniquely specifies the viewing distance. We have now used random dot stereograms in a shape judgement task to show that changes in vertical disparities have no effect on perceived three-dimensional shape. Changes in ocular convergence do alter perceived shape, suggesting substantial changes in the subjects' scaling of horizontal disparities. We conclude that vertical disparities are not used to scale disparities for viewing distance, and that extraretinal signals must be considered when analysing human three-dimensional shape perception.

The horizontal disparity produced by a fixed extent in depth varies roughly inversely with the square of the viewing distance. For a fixed depth to be perceived as invariant (depth constancy), an accurate estimate of viewing distance is required. Viewing distance can in principle be calculated from vertical disparities, as a plot of V/y against x gives a line of slope I/D , where V is the vertical disparity at an image location (x, y) , and I is the interocular separation¹. Bishop recently suggested³ that depth constancy should hold in the presence of natural vertical disparities. If perception of three-dimensional (3-D) shape uses disparity information scaled by an estimate of viewing distance, calculated from vertical disparities, one should expect that changes in the vertical disparity field will alter stereo perception of shape.

So far all of the experimental evidence cited in support of the vertical disparity hypothesis has derived from the 'induced effect'—vertical magnification of one retinal image (in binocular viewing) causes frontoparallel planes to appear slanted⁴. The vertical disparities produced by the magnification are equivalent to those arising from inspection of a truly slanted plane. This situation is also equivalent to viewing a frontoparallel plane with a non-zero angle of gaze. But vertical magnification of one image does not change the slope of the relationship between V/y and x , hence the viewing distance specified by vertical disparities does not change⁵. So theoretically the induced effect should not produce any change in perceived viewing distance, and experimentally none is reported. It is therefore not a direct test of the hypothesis that vertical disparities are used to calculate viewing distance. To clarify the role of vertical disparities in scaling horizontal disparities, we investigated the effect of manipulating vertical disparities on judgements of 3-D shape.

We have previously investigated^{6,7} perception of 3-D shape using a task in which subjects viewed, on a computer monitor, random dot stereograms portraying the surface of a hemi-cylinder emerging from a plane, with the axis of the cylinder aligned horizontally in the frontoparallel plane. At far viewing distances (≥ 107 cm), geometrically circular cylinders were perceived as flattened elliptical cylinders (depth less than half-height), whereas at a near distance (53 cm) they were perceived as elongated (depth greater than half-height)^{6,7}. At an intermediate distance (~ 75 cm), perception was close to veridical. These distortions can readily be accounted for by assuming that subjects mis-estimate viewing distance, and hence apply an inappropriate scaling to the horizontal disparity field.

Subjects were presented with stereograms portraying a series of elliptical cylinders, with varying degrees of elongation in depth. A binary forced choice paradigm was used to determine the elliptical cylinder which appeared circular. This task gives a measure of the angle α subtended by the half-height and the maximum disparity, η , at which the perceived depth, d , is subjectively equal to the perceived half-height, a . Angles α and η are determined by the images and the physical viewing distance, D . It is then possible to calculate a notional viewing distance, D' , at which α and η should correspond to equal physical dimensions, as simple geometry gives the relations

$$a = D' \tan(\alpha) \quad (1)$$

$$\eta = \tan^{-1}(D'/I) - \tan^{-1}((D' - d)/I) \quad (2)$$

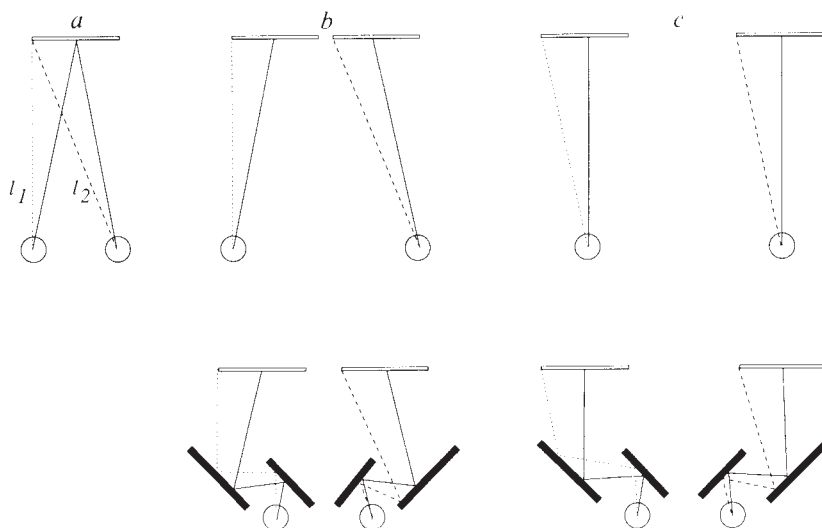
$$\tan(\eta) = Id/(I^2 + D'(D' - d)) \quad (3)$$

where I is the inter-ocular separation. By rearrangement and substitution we obtain

$$D'^2\{\tan(\eta) - \tan(\alpha)\tan(\eta)\} - D'I\tan(\alpha) + I^2\tan(\eta) = 0 \quad (4)$$

This equation is solved for the positive value of D' , which we call the 'scaling distance', as it is a measure of the (usually incorrect) distance estimate which the subject applies implicitly for scaling image data. According to this scheme (similar in principle to that used by Foley⁸), errors in the estimate of D' could also cause mis-estimates of object size.

FIG. 1 Diagram illustrating method for adjusting vertical disparities. The top row illustrates the principle behind the method, and the bottom row illustrates its use in a modified Wheatstone stereoscope. In the stereoscope the vergence angle may be controlled independently of the vertical disparities. The solid line shows the path to the fixation point (centre of the image). *a*, Natural viewing of a plane. Vertical disparities arise because for points away from the midline the pathlengths for rays to the two eyes are different ($l_1 < l_2$). *b*, Stereogram in which natural vertical disparities are preserved. By positioning the centre of each image in the correct position relative to the eye's location, the differences in pathlength are preserved. Naturalistic changes to the vertical disparity field can be made by manipulating these pathlengths. *c*, Stereogram presented so as to remove vertical disparities. The images are positioned so that the pathlength is the same to each side of the plane. By positioning the images at intermediate points, it is possible to produce a pathlength difference which is smaller than in natural viewing. This produces a vertical disparity field corresponding to an increased but finite viewing distance. The proximal pair of mirrors were half-silvered and mounted on rotation stages, and at the start of each experimental run the vergence angle of the eyes, and the centre of each stereo half image, were aligned with a single physical



fixation cross at the correct viewing distance. Rotations of the proximal pair of mirrors were performed in such a way as to preserve the differences in pathlength.

Figure 1 illustrates a method for the manipulation of vertical disparities such that they can be altered to correspond exactly to the vertical disparities of the same object viewed at another viewing distance, while other cues to distance remain unchanged.

Images representing cylinders with a half-height of 5 cm were shown to two subjects (B.G.C. and E.B.J.) at two physical viewing distances, 75 cm and 50 cm. These distances were chosen because in natural viewing the vertical disparities are substantial—up to 8' at 50 cm and 2.5' at 75 cm. These vertical disparities are readily visible, being larger than the most conservative threshold estimate⁹. Figure 2*a* shows that alterations in the vertical disparity field, equivalent to five different viewing distances, had no effect on the perceived shape of the cylinders (there was no change in the scaling of disparity data), at either of the physical distances tested.

This seems to be strong evidence against the hypothesis that

vertical disparities are used to scale horizontal disparity data for viewing distance. This raises the question of what other information (for example, knowledge of convergence) might be used. We manipulated vergence eye position by rotating the front two mirrors of the stereoscope. The images were always positioned so as to maintain a vertical disparity field congruent with the physical viewing distance. As Fig. 2*b* shows, changes in convergence, without changes in vertical disparity, produce sizeable changes in shape perception. We interpret these changes as a rescaling of image data.

The systematic distortions of shape perception with viewing distance, and the changes induced by convergence, leave little doubt that part of the strategy used in this shape judgement task is to scale horizontal disparities by means of an estimate of viewing distance. Knowledge of convergence clearly contributes to this estimate. Nonetheless, it is not the sole contributor—changes in physical viewing distance (dashed line in

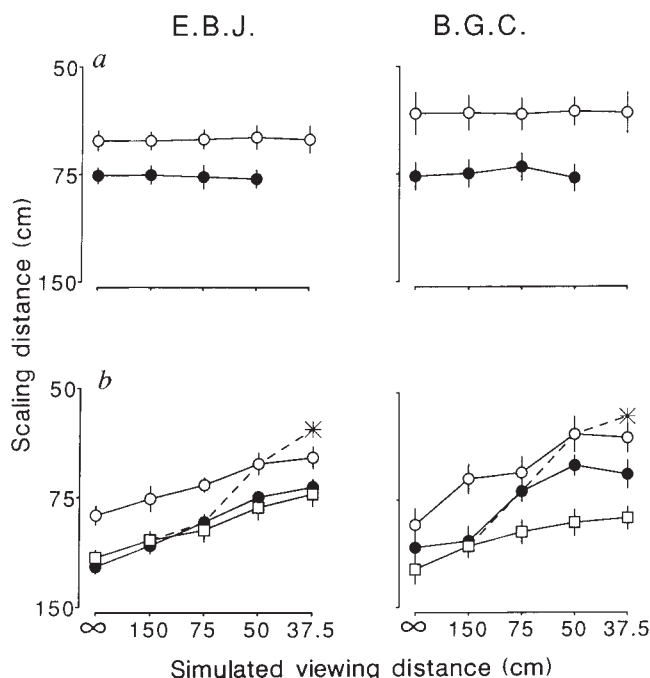


FIG. 2 Effects of independently manipulating vertical disparities (*a*) and convergence angle (*b*) on scaling distance (D' from equation 4). *a*, Scaling distance is plotted against the distance represented by the vertical disparity field. The scale for both axes is reciprocal distance, but the labels show the actual viewing distance for convenience. *b*, Scaling distance is plotted against the distance calculated from the vergence angle. ○, Data collected at a physical viewing distance of 50 cm; ●, 75 cm; □, 150 cm. At a physical distance of 37.5 cm only one data point (shown by the asterisk) was collected, which had with congruent convergence and vertical disparities, for comparison with the vergence manipulations. The dashed lines show the effects of real changes in viewing distance (physically moving the computer monitor in depth) on disparity scaling. Note that the data at different physical viewing distances do not superimpose, showing that cues other than convergence are also used to scale disparities. For each data point five stereograms, representing cylinders of different elongations in depth, were presented in a pseudo-random order for 40 trials each. A cumulative gaussian function was fitted to the forced choice data to yield estimates of the mean and the slope (one standard deviation) of the psychometric function. These estimates were transformed appropriately by equation 4 to give the data and the confidence limits (± 1 s.d.) plotted here. The stereograms were produced by ray-tracing from the nodal point of each eye onto each display pixel, taking account of the display geometry (including the mirror angles). The point at which the rays intersected a 3-D volumetric representation of a cylinder was used to determine the grey-level displayed at each pixel. This technique produces an exact perspective projection of a solid cylinder for each eye, and the use of anti-aliasing allowed disparities to be represented to sub-pixel accuracy.

Fig. 2b) produce larger changes in perceived shape than equivalent changes in convergence alone. Note that the scaling distance is not a veridical estimate of the viewing distance even when the physical viewing distance is altered, in agreement with earlier results⁶⁻⁸. For both subjects, changes in convergence alone alter scaling distance by ~45% of the shift produced by an equivalent real change in viewing distance, or about 25% of the scaling required for complete constancy.

Several schemes have been suggested for the reconstruction of 3-D shape from stereo images by means of purely visual information^{1,2,10,11}. The hypothesis concerning the use of vertical disparities^{1,2} was explicitly tested and shown to be false here. We have also demonstrated that extraretinal cues have a substantial influence on shape perception, an observation that would not be predicted by any purely visual scheme. Future models of human shape perception will need to address questions of sensory-motor integration, in addition to those confined to visual image processing. □

Received 4 June; accepted 6 November 1990.

1. Mayhew, J. E. W. & Longuet-Higgins, H. C. *Nature* **297**, 376-378 (1982).
2. Mayhew, J. E. W. *Perception* **11**, 387-483 (1982).
3. Bishop, P. O. *Proc. R. Soc. B* **237**, 445-469 (1989).
4. Ogle, K. N. *Binocular Vision* (Saunders, Philadelphia, 1950).
5. Frisby, J. *Nature* **307**, 592-593 (1985).
6. Johnston, E. B. *Invest. Ophthalmol. Vis. Sci.* **29**, 399 (1988).
7. Johnston, E. B. *Vision Res.* (in the press).
8. Foley, J. M. *Psychol. Rev.* **87**, 411-434 (1980).
9. Westheimer, G. *Nature* **307**, 632-634 (1984).
10. Koenderink, J. & van Doorn, A. *Biol. Cybernet.* **21**, 755-764 (1976).
11. Rogers, B. & Cagenello, R. *Nature* **339**, 135-137 (1989).

ACKNOWLEDGEMENTS. This work was supported by the MRC, the SERC and the Wellcome Trust.

Two-tone distortion in the basilar membrane of the cochlea

Luis Robles*, Mario A. Ruggero & Nola C. Rich

Department of Otolaryngology, University of Minnesota,
2630 University Avenue S.E., Minneapolis, Minnesota 55414, USA

WHEN humans listen to pairs of tones they hear additional tones, or distortion products, that are not present in the stimulus¹. Two-tone distortion products are also known as combination tones, because their pitches match combinations of the primary frequencies (f_1 and f_2 , $f_2 > f_1$), such as $f_2 - f_1$, $(n+1)f_1 - nf_2$ and $(n+1)f_2 - nf_1$ ($n = 1, 2, 3 \dots$) (refs 2-4). Physiological correlates of the perceived distortion products exist in responses of auditory-nerve fibres⁵⁻⁸ and inner hair cells⁹ and in otoacoustic emissions (sounds generated by the cochlea, recordable at the ear canal)^{7,10-12}. Because the middle ear responds linearly to sound^{13,14} and neural responses to distortion products can be abolished by damage to hair cells at cochlear sites preferentially tuned to the frequencies of the primary tones⁸, it was hypothesized that distortion products are generated at these sites and propagate mechanically along the basilar membrane to the location tuned to the distortion-product frequency^{7,8}. But until now, efforts to confirm this hypothesis have failed^{15,16}. Here we report the use of a new laser-velocimetry technique¹⁷ to demonstrate two-tone distortion in basilar-membrane motion at low and moderate stimulus intensities.

The response of the basilar membrane to two equal-intensity tones was measured at the basal turn of the cochlea in anaesthetized chinchillas. Distortion-product data presented here from a single chinchilla (L17) are representative of similar observations in cochleae from six animals. Figure 1 displays frequency spectra of basilar-membrane responses to primary tones presented at intensities ranging from 60 to 90 dB SPL (decibels 'sound press-

ure level'; that is, referenced to 20 μPa). In addition to peaks at the primary frequencies (7.08 and 7.79 kHz), the spectra include several peaks at distortion-product frequencies both lower and higher than the primary frequencies, such as $2f_1 - f_2$ and $2f_2 - f_1$. The number and amplitude of spectral peaks in the responses vary in a complex manner with the intensity of the two-tone stimuli. The absolute amplitude of some of the peaks actually decreases as the intensity of the primary tones increases from 70 to 90 dB SPL.

Figure 2a illustrates magnitudes of the $2f_1 - f_2$ and $2f_2 - f_1$ distortion products as a function of stimulus intensity. The magnitudes of the distortion products increased with the level of the primary tones up to 80 dB SPL, whereupon they reached a plateau or decreased with further increases in stimulus level. Figure 2a also shows the dependence on intensity of responses to a single tone at the distortion-product frequency (equal to the characteristic frequency, 8.5 kHz). This curve was used to compute the distortion-product 'effective level': the intensity of a single tone at the distortion-product frequency required to

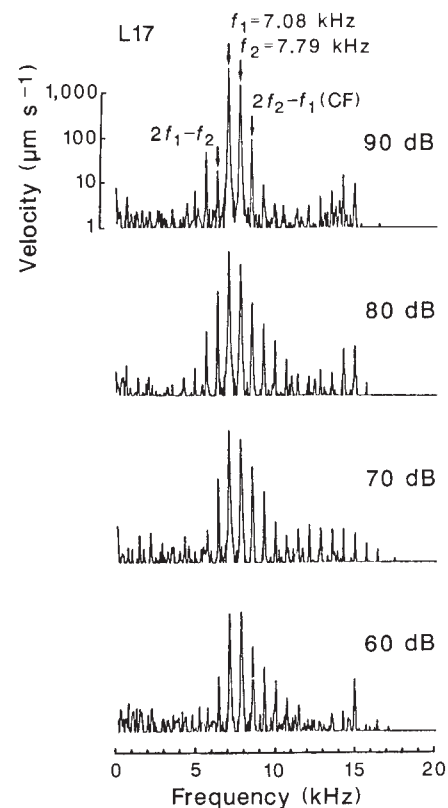


FIG. 1 Frequency spectra of basilar-membrane responses to two-tone stimuli measured using laser velocimetry at the basal turn of the chinchilla cochlea. The spectra were obtained by Fourier transformation of responses to tone pairs presented at 60-90 dB SPL. The primary frequencies were chosen such that $2f_2 - f_1$ coincided with the characteristic frequency (CF, the most sensitive frequency for the cochlear location being studied, in this case 8.5 kHz). The spectra have several peaks corresponding to two-tone distortion products, as well as peaks at the primary frequencies. The spectral peaks corresponding to two prominent distortion products in psychophysical studies ($2f_1 - f_2$ and $2f_2 - f_1$) are indicated by arrows.

METHODS. Basilar-membrane velocity was determined from the Doppler frequency shift of laser light reflected from glass microbeads (10-30- μm diameter) placed on the membrane through a small hole made in the bony shell of the cochlea¹⁷. Mechanical responses were averaged over 1,024 stimulus repetitions and Fourier-transformed to obtain frequency spectra. For two-tone stimulation, tones were delivered independently using two acoustically coupled earphones. This arrangement minimized the generation of intermodulation distortion products. A silver-wire electrode placed on the round window allowed the recording of the whole-nerve action potential to monitor the physiological condition of the cochlea during surgery and data collection.

* Permanent address: Depto. de Fisiología y Biofísica, Fac. de Medicina, Universidad de Chile, Santiago, Chile.

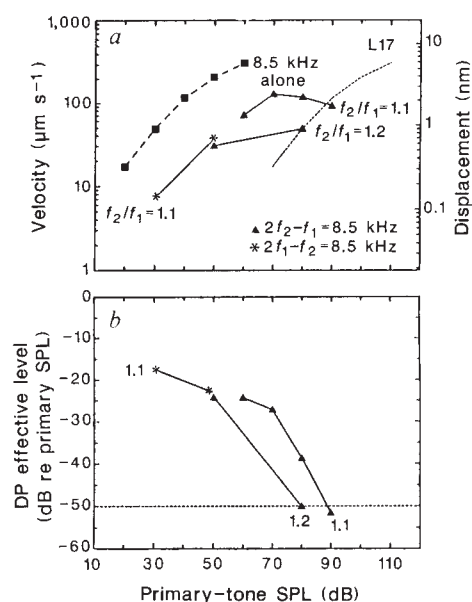


FIG. 2 Magnitudes of the $2f_2 - f_1$ and $2f_1 - f_2$ distortion products as a function of primary-tone intensity (solid lines). The primary-tone frequencies were chosen so that either $2f_2 - f_1$ or $2f_1 - f_2$ coincided with the characteristic frequency (8.5 kHz). *a*, Squares and long-dash line represent the input-output velocity curve for a single tone at the distortion-product frequency. The short-dash line indicates the maximal expected magnitude of artefactual distortion products introduced by the acoustic-stimulus and measuring systems. The $2f_2 - f_1$ data for primary-tone ratio of 1.1 are from the spectra of Fig. 1. *b*, Distortion-product data are plotted as effective level: as intensity of a single tone at the distortion-product frequency required to produce a response of the same magnitude as the distortion product produced by the two-tone stimulus. Effective levels are expressed as decibels relative to primary-tone intensity. The short-dash line indicates that artefactual distortion products were at least 50 dB less than the intensity of the primary tones.

evoke a response of the same magnitude as the distortion product produced by a two-tone stimulus. Distortion-product effective levels (Fig. 2*b*) were highest at the lowest stimulus intensities, ranging between 17 and 24 dB below the level of the primary tones. In other cochleae effective levels of the distortion products were as much as 11 dB less than primary intensities. In most recordings obtained in six cochleae, effective levels of the $2f_1 - f_2$ and $2f_2 - f_1$ distortion products decreased monotonically as stimulus intensity increased. A similar dependence of distortion-product level on stimulus intensity has been described in auditory-nerve responses to $2f_1 - f_2$ distortion products⁵. At or above 80 dB SPL the distortion-product levels were so low that they could be attributed to distortion generated in the stimulus or measuring systems (short-dash lines in Fig. 2).

The magnitudes of the distortion products that we have measured in the chinchilla basilar membrane are comparable to those found in electrophysiological recordings from auditory nerve⁵⁻⁸ of chinchillas and other mammals, and in psychoacoustical studies in humans²⁻⁴. These prominent basilar-membrane distortion products undoubtedly originated in the cochlea, because artefactual distortion products generated in the acoustical and measuring systems were at least 50 dB less than the level of the primary tones (see Fig. 2) and could not have arisen in the middle ear, as its response is linear even at stimulus intensities much higher than used in the present measurements^{13,14}. The previous failure to demonstrate two-tone distortion products in basilar-membrane responses^{15,16} is probably due to inadequate measuring methodology. In particular, the Mössbauer technique, which has provided almost all the evidence on basilar-membrane mechanical nonlinearities, is not well suited to the measurement of responses to complex stimuli, owing to its own nonlinear characteristics. So it is principally as a result of having used an essentially linear technique (laser

velocimetry) that one of the oldest issues in auditory research has been finally settled. Our results, together with other recent basilar-membrane data¹⁷⁻²⁴, suggest that all strongly frequency-dependent nonlinear cochlear phenomena have counterparts in basilar-membrane vibration. Because basilar-membrane responses in damaged cochlea or dead animals are linear^{19,24}, the site of generation of distortion products must be in the living cells of the organ of Corti, probably in the outer hair cells, which almost certainly feed back mechanically to the basilar membrane in normal cochlea^{11,12,23}.

Received 29 August; accepted 30 October 1990.

1. Jones, A. T. *Amer. Phys. Teacher* **3**, 49-51 (1935).
2. Goldstein, J. L. *J. acoust. Soc. Am.* **41**, 676-689 (1967).
3. Smoorenburg, G. F. *J. acoust. Soc. Am.* **52**, 615-632 (1972).
4. Zurek, P. M. & Sachs, R. M. *Science* **205**, 600-602 (1979).
5. Buunen, T. J. F. & Rhode, W. S. *J. acoust. Soc. Am.* **64**, 772-781 (1978).
6. Goldstein, J. L. & Kiang, N. Y. S. *Proc. IEEE* **56**, 981-992 (1968).
7. Kim, D. O., Molnar, C. E. & Matthews, J. W. *J. acoust. Soc. Am.* **67**, 1704-1721 (1980).
8. Siegel, J. H., Kim, D. O. & Molnar, C. E. *J. Neurophysiol.* **47**, 303-328 (1982).
9. Nuttall, A. L. & Dolan, D. F. *J. acoust. Soc. Am.* **87**, 782-790 (1990).
10. Kemp, D. T. *Arch. Otorhinolaryngol.* **224**, 37-45 (1979).
11. Mountain, D. C. *Science* **210**, 71-72 (1980).
12. Siegel, J. H. & Kim, D. O. *Hear. Res.* **6**, 171-182 (1982).
13. Buunen, T. J. F. & Vlamings, M. S. M. G. *J. acoust. Soc. Am.* **69**, 744-750 (1981).
14. Guinan, J. J. Jr & Peake, W. T. *J. acoust. Soc. Am.* **41**, 1237-1261 (1967).
15. Rhode, W. S. in *Psychophysics and Physiology of Hearing* (eds Evans, E. F. & Wilson, J. P.) 27-38 (Academic, London, 1977).
16. Wilson, J. P. & Johnstone, J. R. *Nature* **241**, 206-207 (1973).
17. Ruggero, M. A. & Rich, N. C. *Hear. Res.* (in the press).
18. Patuzzi, R., Sellick, P. M. & Johnstone, B. M. *Hear. Res.* **13**, 19-27 (1984).
19. Robles, L., Ruggero, M. A. & Rich, N. C. *J. acoust. Soc. Am.* **80**, 1364-1374 (1986).
20. Robles, L., Ruggero, M. A. & Rich, N. C. in *Cochlear Mechanisms. Structure, Function and Models* (eds Wilson, J. P. & Kemp, D. T.) 369-375 (Plenum, New York, 1989).
21. Robles, L., Ruggero, M. A. & Rich, N. C. in *Mechanics and Biophysics of Hearing* (eds Dallos, P., Geisler, C. D., Matthews, J. W., Ruggero, M. A. & Steele, C.) (Springer-Verlag, in the press).
22. Nuttall, A. L., Dolan, D. F. & Avinash, G. in *Mechanics and Biophysics of Hearing* (eds Dallos, P., Geisler, C. D., Matthews, J. W., Ruggero, M. A. & Steele, C.) (Springer-Verlag, in the press).
23. Ruggero, M. A. & Rich, N. C. *J. Neurosci.* (in the press).
24. Sellick, P. M., Patuzzi, R. & Johnstone, B. M. *J. acoust. Soc. Am.* **72**, 131-141 (1982).

ACKNOWLEDGEMENTS. This work was supported principally by the NIH and also by the Minnesota Medical Foundation and the Lions 5M Hearing Research Endowment.

Protection of substantia nigra from MPP⁺ neurotoxicity by N-methyl-D-aspartate antagonists

Lechoslaw Turski, Karin Bressler, Klaus-Jürgen Rettig, Peter-Andreas Löschmann & Helmut Wachtel

Research Laboratories of Schering AG, D-1000 Berlin 65, Germany

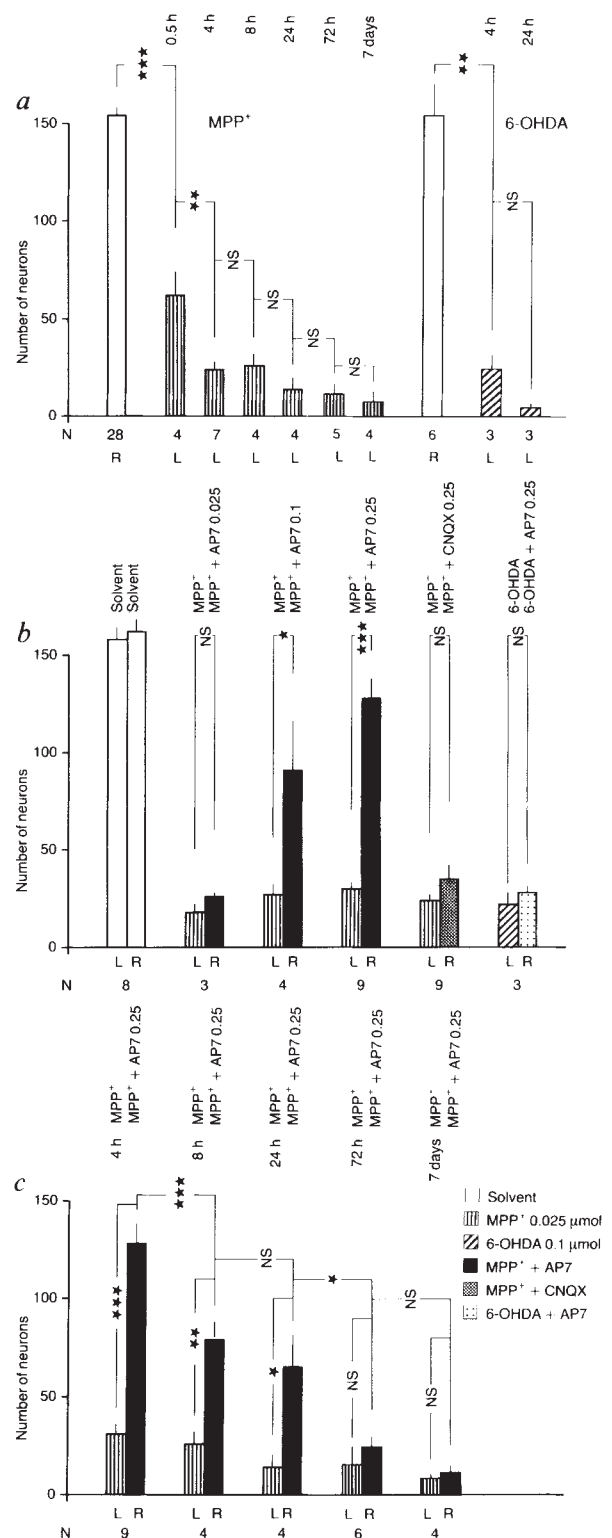
INTAKE of MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) leads to symptoms of Parkinson's disease and produces degeneration of nigrostriatal dopaminergic neurons in humans, giving rise to the hypothesis that this disorder may be caused by endogenous or environmental toxins¹⁻³. Excitation mediated by dicarboxylic amino acids such as L-glutamate or L-aspartate, has been claimed to be involved in pathogenesis of neurodegenerative disorders⁴. We therefore sought to determine whether antagonists active at the NMDA or quisqualate subtypes of L-glutamate receptors prevent toxicity of either MPP⁺ (1-methyl-4-phenyl-pyridinium ion⁵, the active metabolite of MPTP⁶) or the selective dopaminergic neurotoxin 6-OHDA in the rat substantia nigra pars compacta. We report here that certain selective NMDA antagonists (AP7, CPP, MK-801)⁷, but not the preferential quisqualate antagonists CNQX and NBQX⁸, provided short-term (up to 24 h) protection against MPP⁺ toxicity when coadministered into the substantia nigra. Systemic administration of CPP or MK-801 also offered temporary protection for up to 4 h against MPP⁺ toxicity. Repeated systemic administration of either compound prolonged protection against MPP⁺ challenge. Repeated administration for at least 24 h also led to permanent protection, still evident 7 days after intranigral administration of MPP⁺.

FIG. 1 *a*, Time course of neurotoxic action of MPP⁺ (0.025 μ mol) and 6-OHDA (0.1 μ mol) following administration into the substantia nigra in rats. *b*, Dose relationship of protective action of AP7 and CNQX against MPP⁺ and 6-OHDA toxicity in the substantia nigra. *c*, Time course of protection against MPP⁺ toxicity induced by AP7 (0.25 μ mol). Bars represent mean $\pm$ s.e.m. Counts of normal neurons in the substantia nigra pars compacta at the level of AP 2180 (ref. 34) on the right (R) or left (L) side. * P < 0.05; ** P < 0.01; *** P < 0.001, Student's *t*-test.

METHODS. MPP⁺ (1-methyl-4-phenyl-pyridinium iodide; RBI) was dissolved in saline and 6-OHDA (β -(2,4,5-trihydroxyphenyl)-ethylamine HCl; EGA-Chemie) in saline containing 0.1% ascorbic acid. (–)AP7 and CPP (Tocris) were brought into solution with a minimum quantity of NaOH and the final volume made up with saline. MK-801 (Merck Sharp & Dohme) was dissolved in saline, pH was adjusted to 7.35. AP7 (0.025, 0.1 and 0.25 μ mol), CPP (0.25 μ mol) and MK-801 (0.25 μ mol) were coadministered with MPP⁺ (0.025 μ mol) into one substantia nigra. (–)AP7 (0.25 μ mol) was also coadministered into the substantia nigra with 6-OHDA (0.1 μ mol). The contralateral side received MPP⁺ or 6-OHDA alone, and served as drug control. CNQX (Tocris) and NBQX (Novo-Nordisk) was dissolved in NaOH, and the final volume made up with distilled water, pH was adjusted to 7.35. CNQX (0.25 μ mol) and NBQX (0.25 μ mol) were co-administered with MPP⁺ (0.025 μ mol) into one substantia nigra; the contralateral side served as drug control. Drugs were infused in a volume of 1 μ l at a rate of 0.2 μ l min^{–1} into the substantia nigra of pentobarbital-anaesthetized rats (Nembutal, 50 mg kg^{–1} i.p.) at coordinates derived from the stereotaxic atlas of König and Klippel³⁴. The coordinates were: AP 1.9; $\pm$ 1.8; V – 2.1. For morphological examination by light microscopy, rats were anaesthetized with an overdose of pentobarbital and perfused with fixative containing 10% acetic acid, 10% formaldehyde, and 80% methanol. Brains were allowed to fix *in situ* at 4°C for 24 h and were then removed and processed for paraffin embedding. Serial coronal sections of whole brain were cut 10 μ m thick, and every 10th section was mounted on a glass slide and stained with cresyl violet or by Fink and Heimer technique³⁵. For morphological analysis brain sections were visually inspected at 11 different rostro-caudal levels extending from AP 1580 to AP 2580. Quantification of brain damage was performed by direct visual counting of intact neurons by a single observer (L.T.) under blind conditions. The damage in the SNC was quantified at its widest dimension at the level of AP 2180 throughout its mediolateral axis using the Leitz Aristoplan microscope equipped with Leitz counting grid type 513740 at a magnification $\times$ 100. For every SNC, topographical maps of neuronal densities were constructed graphically and total counts were calculated for each side. Normal neurons had typical nuclei with clear nucleoplasm and prominent nucleolus, surrounded by cytoplasm containing Nissl substance. Normal neuron counts were taken for statistical analysis.

To investigate the role of excitatory amino acids in dopaminergic toxicity of pyridine derivatives we used the technique of focal application of MPP⁺ into the rat substantia nigra pars compacta (SNC), the region of the brain containing most dopaminergic cells in this species⁹, and focal or systemic administration of excitatory amino acid antagonists. Intrastriatal administration of MPP⁺ in rats has recently been shown to produce a dose-dependent depletion of striatal dopamine concentration two weeks after treatment³.

To determine the time course of the toxic action of MPP⁺, 28 male Wistar rats, 180–230 g in weight, were microinjected unilaterally with MPP⁺ (0.025 μ mol) into the SNC; the contralateral substantia nigra was infused with vehicle as a control. Rats were killed by transaortic perfusion-fixation 0.5, 4, 8, 24, 72 h and 7 days after intrastriatal treatment, and the brains were serially sectioned and prepared for light microscopy. Six additional rats were used for determination of the time course of the toxic action of 6-OHDA (6-hydroxydopamine) (0.1 μ mol) in the SNC, and were killed 4 and 24 h after intrastriatal infusion of 6-OHDA. Time-course data were obtained from sections through the SNC rostral to the injection site of MPP⁺ or 6-OHDA (Fig. 1*a*). On the vehicle-injected side, 152.8 $\pm$ 3.2 (mean $\pm$ s.e.m.; n = 28) neurons in the SNC were counted. MPP⁺ administration led to a time-dependent reduction in the number of neurons in the SNC. Within 30 min 60% and within 4 h 84% of its neuronal population were lost. Further destruction of the SNC was less dramatic reaching 93% after 7 days. 6-OHDA



administration caused destruction of 85% of neurons after 4 h, with only 4% of the entire neuronal population remaining after 24 h.

Excitatory amino acid antagonists were coadministered into the SNC by infusion with either MPP⁺ or 6-OHDA in 50 rats and their protective action tested after 4 h. Figures 1*b* and 2*a*, *b* show that the NMDA (*N*-methyl-D-aspartate) antagonist AP7 (2-amino-7-phosphonoheptanoate) conferred substantial protection against MPP⁺ toxicity. Its protective effect was dose-

TABLE 1 Protective action of systemic CPP and MK-801 against MPP⁺ toxicity in the rat SNC

Systemic	Treatment	Substantia nigra	Time of observation	Number of neurons mean $\pm$ s.e.m.	N
Dose relationship†					
CPP (0.02 mmol kg ⁻¹)	MPP ⁺ (L)		4h	28.3 $\pm$ 9.4***	4
	Vehicle (R)			147.0 $\pm$ 5.5	
CPP (0.05 mmol kg ⁻¹)	MPP ⁺ (L)		4 h	123.8 $\pm$ 34.4	4
	Vehicle (R)			149.5 $\pm$ 7.7	
CPP (0.1 mmol kg ⁻¹)	MPP ⁺ (L)		4 h	144.4 $\pm$ 10.7	5
	Vehicle (R)			145.4 $\pm$ 6.3	
CPP (1.0 mmol kg ⁻¹)	MPP ⁺ (L)		4 h	140.3 $\pm$ 10.0	6
	Vehicle (R)			145.8 $\pm$ 13.7	
MK-801 (0.002 mmol kg ⁻¹)	MPP ⁺ (L)		4 h	34.0 $\pm$ 7.6**	4
	Vehicle (R)			145.5 $\pm$ 3.4	
MK-801 (0.005 mmol kg ⁻¹)	MPP ⁺ (L)		4 h	103.9 $\pm$ 16.8	3
	Vehicle (R)			152.7 $\pm$ 6.6	
MK-801 (0.01 mmol kg ⁻¹)	MPP ⁺ (L)		4 h	118.1 $\pm$ 27.1	3
	Vehicle (R)			147.0 $\pm$ 2.0	
Time course					
CPP (0.1 mmol kg ⁻¹)	MPP ⁺ (L)		24 h	14.5 $\pm$ 5.3***	4
	Vehicle (R)			149.5 $\pm$ 3.4	
CPP (0.1 mmol kg ⁻¹)	MPP ⁺ (L)		72 h	26.6 $\pm$ 8.2***	7
	Vehicle (R)			142.3 $\pm$ 7.2	
MK-801 (0.01 mmol kg ⁻¹)	MPP ⁺ (L)		24 h	49.3 $\pm$ 26.3*	4
	Vehicle (R)			142.3 $\pm$ 7.2	
MK-801 (0.01 mmol kg ⁻¹)	MPP ⁺ (L)		72 h	21.7 $\pm$ 7.7*	3
	Vehicle (R)			147.5 $\pm$ 6.4	
Repeated administration‡					
(i) 6 times					
Vehicle	MPP ⁺ (L)		24 h	39.7 $\pm$ 15.7*	3
	Vehicle (R)			141.7 $\pm$ 10.0	
CPP (0.1 mmol kg ⁻¹)	MPP ⁺ (L)		24 h	143.4 $\pm$ 14.4	5
	Vehicle (R)			151.8 $\pm$ 4.5	
MK-801 (0.01 mmol kg ⁻¹)	MPP ⁺ (L)		24 h	123.4 $\pm$ 20.3	5
	Vehicle (R)			148.2 $\pm$ 4.0	
Vehicle	MPP ⁺ (L)		7 days	19.8 $\pm$ 4.2***	4
	Vehicle (R)			146.8 $\pm$ 3.1	
CPP (0.1 mmol kg ⁻¹)	MPP ⁺ (L)		7 days	135.3 $\pm$ 9.2	6
	Vehicle (R)			146.7 $\pm$ 3.7	
MK-801 (0.01 mmol kg ⁻¹)	MPP ⁺ (L)		7 days	117.6 $\pm$ 14.4	5
	Vehicle (R)			142.8 $\pm$ 2.3	
(ii) 18 times					
Vehicle	MPP ⁺ (L)		72 h	32.8 $\pm$ 8.6***	5
	Vehicle (R)			144.0 $\pm$ 3.4	
CPP (0.1 mmol kg ⁻¹)	MPP ⁺ (L)		72 h	115.2 $\pm$ 16.9	5
	Vehicle (R)			144.6 $\pm$ 3.9	
MK-801 (0.01 mmol kg ⁻¹)	MPP ⁺ (L)		72 h	141.0 $\pm$ 6.0	4
	Vehicle (R)			146.5 $\pm$ 1.6	
Vehicle	MPP ⁺ (L)		7 days	26.8 $\pm$ 3.2***	4
	Vehicle (R)			144.3 $\pm$ 2.4	
CPP (0.1 mmol kg ⁻¹)	MPP ⁺ (L)		7 days	121.8 $\pm$ 12.9	6
	Vehicle (R)			146.3 $\pm$ 4.2	
MK-801 (0.01 mmol kg ⁻¹)	MPP ⁺ (L)		7 days	138.8 $\pm$ 14.9	5
	Vehicle (R)			149.8 $\pm$ 3.9	

N, number of animals. R, right side. L, left side.

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ vs. respective vehicle controls, Student's *t*-test.

† CPP and MK-801 were administered i.p. 30 min before microinjection of MPP⁺ (0.025 μ mol) into one substantia nigra; the contralateral side was infused with vehicle and served as a control. Brains were analysed morphometrically 4, 24 and 72 h after microinjection of MPP⁺.

‡ CPP, MK-801 and vehicle were administered every 4 h, either 6 or 18 times. The first systemic injection of CPP, MK-801 or vehicle was given 30 min before the microinjection of MPP⁺ (0.025 μ mol) into one substantia nigra; the contralateral side was injected with vehicle and served as a control. The brains were analysed morphometrically 24, 72 h and 7 days after microinjection of MPP⁺.

dependent in the range 0.025–0.25 μ mol, and was also time-dependent. Protection observed at 4 h was more pronounced than that found after 8 or 24 h (Fig. 1c). No protection by AP7 against MPP⁺ toxicity in the SNC was detected 72 h and 7 days after its coadministration (Fig. 1c). Similar protection against MPP⁺ toxicity was seen after coadministration into the SNC of 0.25 μ mol CPP (3-(($\pm$)-2-carboxypiperazin-4-yl)-propyl-1-phosphonate) (142.3 $\pm$ 10.6 neurons on CPP+MPP⁺ site versus 36.7 $\pm$ 7.3 neurons on MPP⁺ site; $n = 3$, $P < 0.001$, Student's *t*-test) and 0.25 μ mol MK-801 ((+)-5-methyl-10,11-dihydro-5H-dibenzo-(a,d)cycloheptan-5,10-imine maleate) (106.3 $\pm$ 19.9

neurons on MK-801+MPP⁺ site versus 43.3 $\pm$ 13.4 neurons on MPP⁺ site; $n = 6$, $P < 0.05$). The quisqualate antagonists CNQX (6-cyano-7-nitroquinoxaline-2,3-dione) (Fig. 1b) and NBQX (2,3-dihydroxy-6-nitro-7-sulphamoyl-benzo(F)quinoxaline) (33.0 $\pm$ 5.5 neurons on NBQX+MPP⁺ site versus 24.0 $\pm$ 5.2 neurons on MPP⁺ site; $n = 5$, non-significant) exerted no protective effect against MPP⁺ toxicity in the SNC. Similarly, AP7 conferred no protection against 6-OHDA toxicity (Figs 1b and 2c, d).

CPP and MK-801 are NMDA antagonists which penetrate the blood-brain barrier, and were therefore used to ascertain

whether there was a protective action against MPP⁺ toxicity on systemic administration. The drugs were administered intraperitoneally (i.p.) 30 min before intranigral infusions of MPP⁺ (0.025 μ mol); the contralateral substantia nigra was infused with vehicle and served as a control. CPP (0.02–0.1 mmol kg⁻¹) and MK-801 (0.002–0.01 mmol kg⁻¹) provided dose-dependent protection against MPP⁺ toxicity in the SNC (Table 1). The protective action of both CPP (0.1 mmol kg⁻¹) and MK-801 (0.01 mmol kg⁻¹) was time-dependent and was seen 4 h but not 24 or 72 h after intranigral administration of the toxin (Table 1). Therefore, we also tested whether repeated administration of the NMDA antagonists could prolong protection against MPP⁺ toxicity. Table 1 and Fig. 2*e,f* show that repeated administration of CPP (0.1 mmol kg⁻¹) or MK-801 (0.01 mmol kg⁻¹) every 4 h for 24 h or 72 h offered efficient protection against MPP⁺ toxicity in the SNC. This protective effect remained permanent and was still evident 7 days after intranigral administration of MPP⁺ (Table 1).

These data show that temporary protection against MPP⁺ toxicity in the substantia nigra is offered by the competitive NMDA antagonists AP7 and CPP and the non-competitive NMDA antagonist MK-801. Repeated systemic administration of either CPP or MK-801 prolongs protection against MPP⁺ toxicity, and when continued for 24 h, the protective effect becomes permanent. Such observations contrast with those made with systemic administration of MK-801 by Sonsalla and her coworkers¹⁰ who found that MK-801 protected against dopaminergic toxicity of methamphetamine in mice but not against that induced by MPTP. There may be several reasons for this disparity. First, we used the technique of focal application of toxins, whereas others¹⁰ have used systemic administration. Second, we found a temporary protection against MPP⁺ toxicity; this effect was detectable for up to 24 h after focal and for up to 4 h after systemic administration of the NMDA antagonists. Sonsalla and colleagues examined their animals for the first time 3 days after administration of MPTP¹⁰, which may mean that a possible temporary protection was overlooked. Third, the use of MPP⁺ rather than MPTP and/or species differences (rats vs mice) may also have contributed to the differences observed.

Another important aspect of our work is that no protection against MPP⁺ toxicity was conferred by the quisqualate antagonists CNQX and NBQX. Although the role of non-NMDA receptors in neurodegenerative disorders is not yet clear, initial experimental evidence indicates that quisqualate receptors are involved in ischaemia and ageing^{4,8}. Our observation may therefore indicate that quisqualate-dependent mechanisms do not contribute to the early toxicity of MPP⁺. Similarly, the neurotoxic action of 6-OHDA was not affected by AP7, suggesting that NMDA-mediated excitation does not contribute to the toxic action of this compound.

Two other issues are addressed by the data presented above. We have shown for the first time that NMDA antagonists may protect against MPP⁺ toxicity. This effect is temporary but may be prolonged by repeated administration of NMDA antagonists and finally may become permanent. The question arises of the mechanisms responsible for such profiles of protective action and their impact in the prevention of Parkinson's disease or related neurodegenerative disorders. Studies with mitochondrial preparations have shown that MPP⁺ interferes with energy metabolism and depletes cell energy supplies^{11,12}. Accumulation of MPP⁺ and depletion of energy resources are regarded as two essential steps leading to cell death¹³. One consequence of such energy deprivation in neurons may be an alteration in cell membrane function leading to partial depolarization and hence to a relief of the voltage-dependent Mg²⁺ block of NMDA receptor channels¹⁴. Maintenance of the resting membrane potential is dependent on glucose metabolism, ATP synthesis and functioning Na⁺, K⁺-ATPases¹⁵. Relief of the Mg²⁺ block enables excitatory amino acids to act persistently at their recep-

tors, resulting in the opening of ion channels, and consequently enables glutamate to become neurotoxic¹⁵. It has indeed been suggested that the NMDA receptor mediates glutamate neurotoxicity when intracellular energy metabolism is reduced^{15,16}. It seems possible that similar mechanisms are responsible at least in part for the toxic action of MPP⁺, but not that of 6-OHDA, in the substantia nigra.

A remaining open question concerns the temporary nature of the protection offered by NMDA antagonists against MPP⁺ toxicity. It may simply be that MPP⁺ remains in the brain for longer than the NMDA antagonists and thus the toxic effect of MPP⁺ overcomes the protection provided by them. Several lines of evidence support this inference^{17–22}. MPP⁺ can be detected in the rodent brain for at least 24 h after intracerebral administration¹⁷; similar results follow systemic administration of MPTP^{18,19}. In contrast, NMDA antagonists, such as AP7 or MK-801, are detected in the rodent brain or cerebrospinal fluid only within 3–4 h after their systemic administration^{20–22}. Alternatively, MPP⁺ may not be neurotoxic by itself, but induce irreversible changes in neuronal function, which need a step mediated by excitatory amino acids for toxicity to be expressed. In that case, protection by NMDA antagonists should be time-limited and the toxic action of MPP⁺ nonselective. Both these

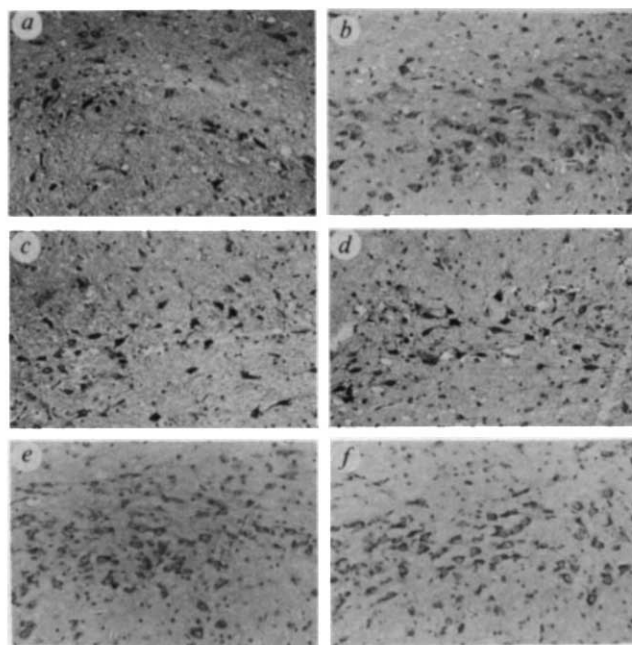


FIG. 2 Protection from neuropathological sequelae of MPP⁺ or 6-OHDA in the rat substantia nigra after co-administration of AP7 or systemic administration of CPP. *a*, SNC with grave breakdown of the cell structure, massive oedematous changes and extensive disintegration of neurophil 4 h after intranigral microinjection of MPP⁺ (0.025 μ mol). Similar type of morphological change is also seen in the pars reticulata of the substantia nigra (SNR). *b* (side contralateral to *a*), Normal cytoarchitecture of the SNC 4 h after treatment with MPP⁺ and AP7 (0.25 μ mol). No signs of degeneration are evident in the SNR. *c*, Destruction of the SNC in a rat microinjected with 6-OHDA (0.1 μ mol), observed 4 h after treatment. *d* (side contralateral to *c*), No visible signs of protection against 6-OHDA toxicity in SNC after co-administration of AP7 (0.25 μ mol), apparent 4 h after treatment. The extent of pathological tissue reaction is similar to that seen in *c*. *e*, Normal morphology of the SNC in a rat subjected to intranigral microinjection of vehicle and 6 intraperitoneal (i.p.) injections of CPP (0.1 mmol kg⁻¹) (6 $\times$ every 4 h). *f* (side contralateral to *e*), No sign of pathological reaction to MPP⁺ seen in the SNC after 6 i.p. injections of CPP (0.1 mmol kg⁻¹), given as in *e*. Survival time in *e* and *f*, 24 h. Cresyl violet stain. Magnifications $\times 100$.

features seem to apply to the characteristics of MPP⁺ toxicity in brain and cell culture²³⁻²⁵.

The protective action of NMDA antagonists against MPP⁺ toxicity is additional evidence for the hypothesis that neurotoxicity mediated by excitatory amino acids may be involved in the pathogenesis of Parkinson's disease²⁶. If this is true, these observations form a new basis for therapeutic strategies directed towards prevention of Parkinson's disease. That such a hypothesis may have broader importance is suggested by observations linking amyotrophic lateral sclerosis-parkinsonism-dementia complex of Guam and excitatory amino acids of plant origin²⁷ or parkinsonism associated with spontaneous degeneration of nigrostriatal neurons and physical trauma in boxers with long careers (dementia pugilistica)²⁸. An unusual amino acid, β -N-methylamino-L-alanine, is believed to be the cause of the Guam disease complex²⁷. It has recently been shown that this compound acts by activation of the glutamate receptors²⁹. The effects of concussive brain and spinal cord injury are associated with raised extracellular concentrations of endogenous excitatory amino acids and may be prevented by NMDA antagonists³⁰. It is therefore reasonable to suggest that nigrostriatal degeneration and subsequent parkinsonism observed in dementia pugilistica may be due to the unleashed toxic action of endogenous excitatory amino acids on dopaminergic cells in the substantia nigra²⁸. Another link between parkinsonism and excitatory amino acids is suggested by the unexpected finding that anticholinergic drugs and amantadine derivatives, long used to treat this disease, are potent NMDA antagonists^{31,32}. We do not yet know whether the early use of anticholinergics or amantadine derivatives could postpone indications for antiparkinsonian medication with L-DOPA and could possibly delay the progression of Parkinson's disease as does treatment with monoamine oxidase B inhibitors³³. Should that be the case, then a new strategy, based on NMDA antagonism, for the neuroprotective or preventive therapy of Parkinson's disease and perhaps other neurodegenerative disorders could be offered. □

NMDA receptor agonists selectively block N-type calcium channels in hippocampal neurons

Natalya I. Chernevskaya, Alexander G. Obukhov
& Oleg A. Krishtal*

Bogomoletz Institute of Physiology, 252024, Kiev,
Bogomoletz str. 4, USSR

THE modulation of voltage-dependent calcium channels by various neurotransmitters has been demonstrated in many neurons¹⁻⁴. Because of the critical role of Ca²⁺ in transmitter release and, more generally, in transmembrane signalling, this modulation has important functional implications. Hippocampal neurons possess low-threshold (T-type) Ca²⁺ channels and both L- and N-type high voltage-activated Ca²⁺ channels.⁵⁻⁷ N-type Ca²⁺ channels are blocked selectively by ω -conotoxin^{8,9} and adenosine^{10,11}. These substances both block excitatory synaptic transmission in the hippocampus¹²⁻¹³, whereas dihydropyridines, which selectively block L-type channels¹⁴, are ineffective¹². Excitatory synaptic transmission in the hippocampus displays a number of plasticity phenomena that are initiated by Ca²⁺ entry through ionic channels operated by N-methyl-D-aspartate (NMDA) receptors^{15,16}. Here we report that NMDA receptor agonists selectively and effectively depress N-type Ca²⁺ channels which are involved in neurotransmitter release from presynaptic sites. The inhibitory effect is eliminated by the competitive NMDA antagonist D-2-amino-5-phosphonovalerate, does not require Ca²⁺ entry into the cell, and is probably receptor-mediated. This phenomenon may provide a negative feedback between the liberation of excitatory transmitter and entry of Ca²⁺ into the cell, and could be important in presynaptic inhibition and in the regulation of synaptic plasticity.

Hippocampal pyramidal neurons acutely isolated from the CA1 and CA3 region have at least three types of Ca²⁺ channels. Here we demonstrate that the same component of Ca²⁺ current that is reversibly blocked by adenosine can also be blocked by excitatory amino acids (EAAs). An example is shown in Fig. 1a in which we use EAA glutamate. The left panel demonstrates the blocking action of adenosine. This block is completely reversible, as shown in the control trace (1) of the middle panel. After washout, Glu (20 μ M) blocks the Ca²⁺ current to the same extent as adenosine (middle panel). Adenosine has no additional effect (right panel). The Glu block is largely irreversible within the time limitations of our experiments (currents usually ran down within 10-20 min). This apparent irreversibility is probably a result of some experimental feature such as intracellular perfusion. In a separate series of experiments, hippocampal slices were pretreated with Glu (20 μ M) for 15 min and then washed during 15-60 min. Neurons isolated from these slices were fully sensitive to subsequent application of EAA or adenosine.

The EAA-sensitive current is blocked by the irreversible N-type Ca²⁺ channel blocker ω -conotoxin. Neurons isolated from slices pretreated with ω -conotoxin for 30-60 min were insensitive to EAA or adenosine (total $n = 11$; data not shown). The long pretreatment was necessary for the complete development of the ω -conotoxin blocking action¹².

The current inhibited by EAA seems to correspond well to the N-type Ca²⁺ current. The contribution of this N-current to the total Ca²⁺ current is largest at depolarizations between -30 and -10 mV. Current-voltage relationships obtained in different cells (Fig. 1b) verify this, as well as the lack of EAA effect on T- and L-type currents, which is apparent in the case of the T current from a cell in which high voltage-activated Ca²⁺ channels

Received 31 August; accepted 5 November 1990.

- Langston, J. W., Ballard, P., Tetrad, J. W. & Irwin, I. *Science* **219**, 979-980 (1983).
- Burns, R. S. et al. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4546-4550 (1983).
- Bradbury, A. J. et al. *Nature* **319**, 56-57 (1986).
- Olney, J. W. In *Frontiers in Excitatory Amino Acid Research* (eds Cavalheiro, E., Lehmann, J. & Turski, L.) 589-596 (Liss, New York, 1988).
- Markey, S. P., Johannesssen, J. N., Chiueh, C. C., Burns, R. S. & Herkenham, M. A. *Nature* **311**, 464-467 (1984).
- Heikkila, R. E., Manzino, L., Cabbat, F. S. & Duvoisin, R. C. *Nature* **311**, 467-469 (1984).
- Olvermann, H. J., Jones, A. W. & Watkins, J. C. *Nature* **307**, 460-462 (1984).
- Sheardown, M. J., Nielsen, E. O., Hansen, A. J., Jacobsen, P. & Honore, T. *Science* **247**, 571-574 (1990).
- Fallon, J. H. & Loughlin, S. E. In *The Rat Nervous System* (ed. Paxinos, G.) 353-403 (Academic, Sydney, 1985).
- Sonsalla, P. K., Nicklas, W. J. & Heikkila, R. E. *Science* **243**, 398-400 (1989).
- Nicklas, W. J., Vyas, I. & Heikkila, R. E. *Life Sci.* **36**, 2503-2508 (1985).
- Ramsay, R. R., Salach, J. I. & Singer, T. P. *Biochem. biophys. Res. Commun.* **134**, 743-748 (1986).
- Di Monte, D. & Smith, M. T. *Rev. Neurosci.* **2**, 67-82 (1988).
- Nowak, L., Bregestovskii, P., Ascher, P., Herbet, A. & Prochiantz, A. *Nature* **307**, 462-465 (1984).
- Novelli, A., Reilly, J. A., Lyso, P. G. & Henneberry, R. C. *Brain Res.* **451**, 205-212 (1988).
- Rothman, S. *Science* **220**, 536-538 (1983).
- Irwin, I., DeLanney, L. E., Di Monte, D. & Langston, J. W. *Neurosci. Lett.* **101**, 83-88 (1989).
- Johannessen, J. N., Chiueh, C. C., Burns, R. S. & Markey, S. P. *Life Sci.* **36**, 219-224 (1985).
- Ohya, Y. et al. *Neurosci. Lett.* **105**, 221-226 (1989).
- Chapman, A., Collins, J. F., Meldrum, B. S. & Westerberg, E. *Neurosci. Lett.* **37**, 75-80 (1983).
- Compton, R. P., Hood, W. F. & Monahan, J. B. *Neurosci. Lett.* **84**, 339-344 (1988).
- Vezzani, A. et al. *J. Pharmacol. exp. Ther.* **249**, 278-283 (1989).
- Gibb, W. R. G., Costall, B., Domeney, A. M., Kelly, M. E. & Naylor, R. J. *Brain Res.* **461**, 361-366 (1988).
- Sun, C. J. et al. *J. Neural Transm.* **74**, 75-86 (1988).
- Christie-Pope, B. C., Burns, R. S. & Whetsell, W. O. *Exp. Neurol.* **104**, 235-240 (1989).
- Caine, D. B., Eisen, A., McGeer, E. & Spencer, P. *Lancet* **ii**, 1067-1070 (1986).
- Spencer, P. S. et al. *Science* **237**, 517-522 (1987).
- Mawdsley, C. & Ferguson, F. R. *Lancet* **ii**, 795-801 (1983).
- Weiss, J. H. & Choi, D. W. *Science* **241**, 973-975 (1988).
- Faden, A. I., Demeduk, P., Panter, S. S. & Vink, R. *Science* **244**, 798-800 (1989).
- Olney, J. W. et al. *Eur. J. Pharmacol.* **142**, 319-320 (1987).
- Bormann, J. *Eur. J. Pharmacol.* **166**, 591-592 (1989).
- Tetrad, J. W. & Langston, J. W. *Science* **245**, 519-522 (1989).
- König, J. F. R. & Klippel, R. A. *The Rat Brain: A Stereotaxic Atlas of the Forebrain and Lower Parts of the Brain Stem* (William and Wilkins, Baltimore, 1963).
- Fink, R. P. & Heimer, L. *Brain Res.* **4**, 369-374 (1967).

* To whom correspondence should be addressed.

ran down with time (left panel). Another cell with minimal L current is used to demonstrate effects on N-current (middle panel). L-type current is isolated by inactivating the T- and N-currents with positive steady-state holding potential ($V_h = -50$ mV). EAA had no effect on this component (right panel). This is also demonstrated in a single cell where individual traces were obtained from $V_h = -50$ and -90 mV in control solutions and after application of Glu (Fig. 1c). The trace obtained at -90 mV contains a large decaying component which is largely

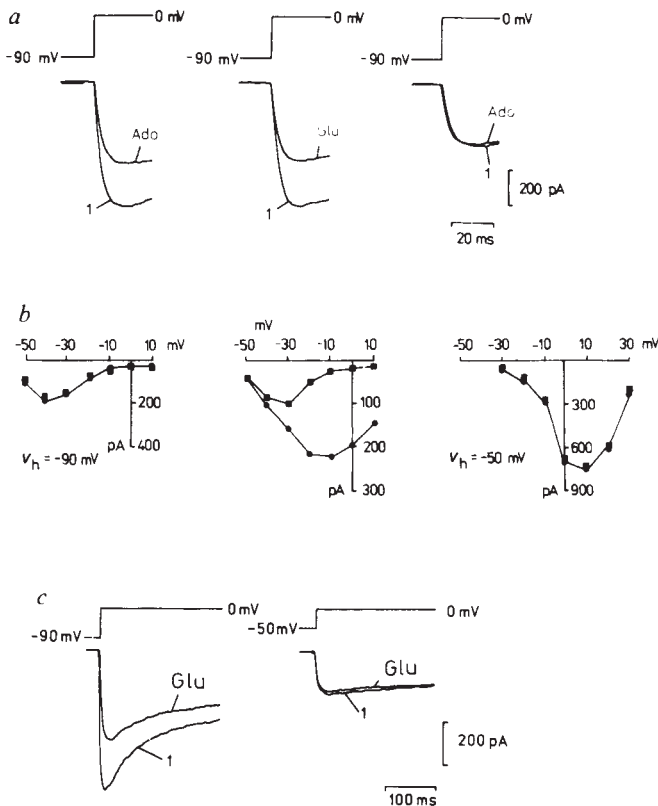


FIG. 1 Glu inhibition of adenosine (Ado)-sensitive Ca^{2+} currents in hippocampal pyramidal neurons. *a*, Same neuron in all panels from left to right; inhibition of Ca^{2+} current by $10 \mu\text{M}$ adenosine (trace compared with control trace 1); inhibition of Ca^{2+} current by $20 \mu\text{M}$ Glu (control trace 1 was obtained after washout of adenosine); the lack of adenosine ($10 \mu\text{M}$) effect on the Ca^{2+} current remaining after the action of Glu. Currents were recorded 1 min after applications with the last trace before the application trace used as a control. These effects were seen in essentially all neurons ($n = 7$). *b*, Peak Ca^{2+} current I - V relationships obtained in control saline (circles) and in the presence of $20 \mu\text{M}$ Glu (squares). Left panel: measurements began 30 min after initiating intracellular perfusion so that only T-current remained and demonstrated a total lack of sensitivity to Glu; $V_h = -90$ mV (holding potential). Middle: measurements started within 5 min after breaking into the cell (a cell with almost no L-current was selected). $V_h = -90$ mV. Right: the lack of Glu effect ($20 \mu\text{M}$) on L-current. The N-current is inactivated by shifting V_h to -50 mV. The successive depolarizing steps were applied at 0.5 min intervals in all experiments. *c*, Effect of Glu on the decaying and sustained components of Ca^{2+} current. Control traces were obtained from $V_h = -90$ mV and -50 mV (1). The protocol was repeated after addition of $20 \mu\text{M}$ Glu. Glu has no effect on the sustained current elicited from $V_h = -50$ mV, but diminishes the more rapidly decaying component seen at $V_h = -90$ mV.

METHODS. 400- μm -thick hippocampal slices of 12-14 day old Wistar rats were incubated at 30°C in Dulbecco's phosphate-buffered saline solution (Sigma) with 0.5 mM MgCl_2 containing 0.5% trypsin (Serva; 1:250) for 40-50 min²⁵. Single cells were released by trituration shortly before the experiments. Control medium contained (in mM): 140 N -methyl-D-glucamine chloride, 5 CsCl , 10 HEPES-CsOH , 10 glucose , 20 TEA-Cl , 10 CaCl_2 , pH 7.4. Internal solution: 100 N -methyl-D-glucamine phosphate, 20 TEA-Cl , 20 Tris-HCl , 10 EGTA , 2 Mg-ATP , 0.01 cyclicAMP (sodium salt), 1 GTP (sodium salt), pH 7.3. Conventional whole-cell voltage clamp was used in combination with rapid application technique²⁶. Experiments were carried out at $21-23^\circ\text{C}$.

TABLE 1 Comparison of the ability of substances to inhibit (+) high-threshold Ca^{2+} currents and excitatory synaptic transmission in hippocampal areas CA1 and CA3

	N-current	L-current	EST
DHP	-	+ (refs 22, 23)	- (ref. 12)
ω -CgTX	+	- (ref. 12)	+ (refs 12, 13)
Ado	+	- (refs 11, 24)	+ (refs 13, 24)
NMDA-agonists	+	-	+ (ref. 19)

Abbreviations: EST, excitatory synaptic transmission; DHP, dihydropyridine; ω -CgTX, ω -conotoxin; Ado, adenosine.

blocked by Glu, whereas there is little effect on the sustained current obtained at -50 mV.

Table 1 demonstrates a complete correlation between the pharmacological properties of hippocampal excitatory synaptic transmission and N-current, indicating that the latter may have a key role in triggering transmitter release in this system of excitatory synapses.

The participation of NMDA receptors in blocking N-type currents is indicated by pharmacological experiments. Both of the endogenous EAAs Glu and Asp block the current completely at concentrations as low as $2 \mu\text{M}$ for Glu (lower concentrations were not tested). NMDA was also effective (Fig. 2a). There was no block in the presence of the competitive NMDA-receptor antagonist D-2-amino-5-phosphonovalerate¹⁷ (Fig. 2b), whereas the selective non-NMDA agonists, kainate ($20 \mu\text{M}$) and (*R,S*)- α -amino-3-hydroxy-5-methylisoxazole-propionic acid¹⁸ ($20 \mu\text{M}$) had no effect. These pharmacological results account for the inhibition of excitatory synaptic transmission not only by Glu, but equally by Asp¹⁹, which does not activate non-NMDA EAA receptors that are normally responsible for hippocampal low-frequency synaptic transmission²⁰.

Kynurenic acid prevents the opening of NMDA receptor-operated channels by interacting with the glycine modulatory receptor site²¹. In the absence of measurable quantities of Gly and in the presence of kynurenic acid, there is no detectable

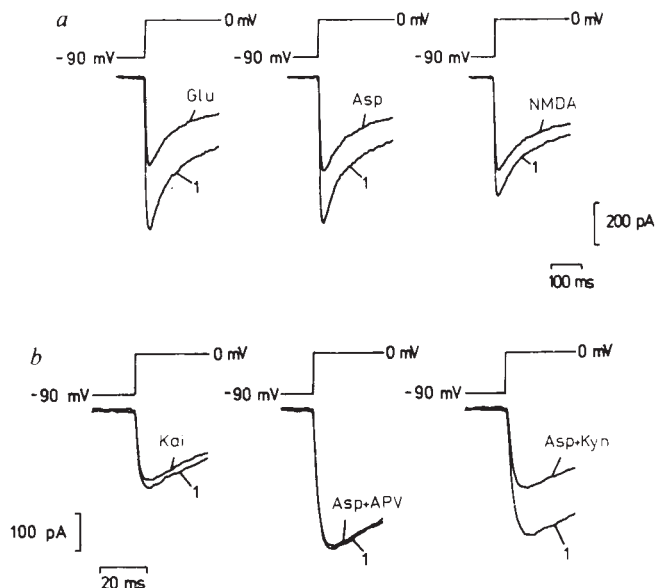


FIG. 2 Pharmacology of EAA block of Ca^{2+} currents. *a*, Effects of $20 \mu\text{M}$ Glu, $100 \mu\text{M}$ Asp and $20 \mu\text{M}$ NMDA. Trace 1 is the control trace. Ca^{2+} current were elicited by depolarization to 0 mV , $V_h = -90 \text{ mV}$. *b*, Left, $20 \mu\text{M}$ kainic acid does not block Ca^{2+} current; middle, $500 \mu\text{M}$ D-2-amino-5-phosphonovalerate (APV) blocks the action of Asp ($500 \mu\text{M}$); right, $1 \mu\text{M}$ kynurenic (Kyn) acid does not protect Ca^{2+} current from blocking action of $500 \mu\text{M}$ Asp.

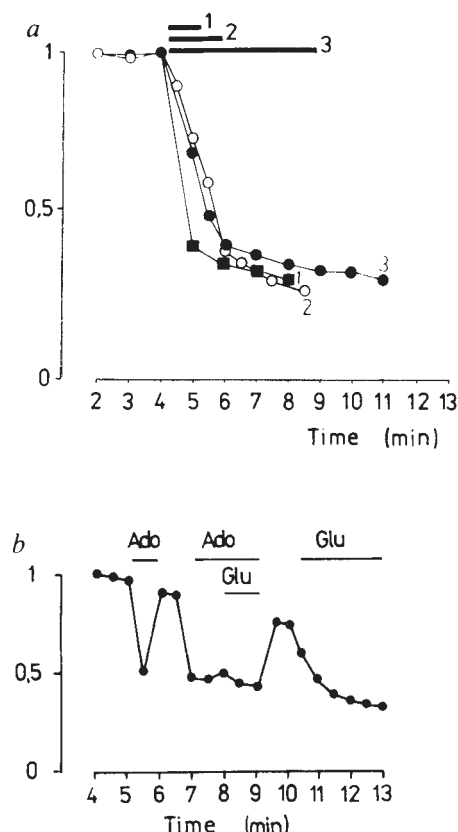


FIG. 3 Adenosine protection of Glu 'irreversibility'. *a*, Relative Ca^{2+} currents (I/I_{max}) in three separate cells. Bars correspond to Glu ($20 \mu\text{M}$) application for 1 (cell 1), 2 (cell 2) or 5 min (cell 3). Several minutes wash does not reverse the block. *b*, Adenosine (Ado) ($10 \mu\text{M}$) block readily reverses. Application of Glu ($20 \mu\text{M}$) concurrently with adenosine does not produce additional block, but readily reverses. The currents are still susceptible to Glu block after the wash.

current from NMDA receptor-operated channels. EAA inhibits N-current in the presence of kynurenic acid ($1 \mu\text{M}$; Fig. 2b). EAA block was equally effective if Ba^{2+} ions replaced Ca^{2+} as the charge carriers. Thus the EAA blocking effect does not depend on the entry of substantial amounts of Ca^{2+} by NMDA-operated channels. Therefore EAA block may be mediated by direct NMDA receptor action through some effector other than entry of Ca^{2+} through the channel. But our experiments do not exclude intracellular changes in Ca^{2+} concentration, which may be needed in extremely small amounts.

As mentioned above, adenosine block is easily reversible (Fig. 1a and 3b); Glu block is usually irreversible under our experimental conditions (Fig. 3a). But if Glu is applied to cells exposed to adenosine at the same time, the total block is largely reversible (Fig. 3b). This protection from Glu irreversibility indicates a tight link between the actions of adenosine and EAA on N-type channels.

We have shown previously that in the presence of Asp or Glu, excitatory synaptic transmission between CA3 and CA1 neurons becomes insensitive to adenosine and ω -conotoxin¹³. This reversible effect is accompanied by a long-lasting change in excitatory synaptic transmission pharmacology and may be related to plasticity¹³. Now we have found that adenosine reversibly protects the N-current from the blocking action of EAA (Fig. 3b). We suggest that a regulatory complex exists which controls N-channels through functionally interconnected NMDA and adenosine receptors. Whether this regulatory complex operates through physical interconnections or through networks of second messenger systems awaits further experi-

ment. But, there are clear implications for the presynaptic regulation of transmitter release. □

Received 3 October; accepted 9 November 1990.

1. Dunlap, K. & Fischbach, G. D. *J. Physiol., Lond.* **317**, 519–535 (1981).
2. Deisz, R. A. & Lux, H. D. *Neurosci. Lett.* **56**, 205–210 (1985).
3. Bean, B. P. *Nature* **340**, 153–156 (1989).
4. Toselli, M., Lang, J., Costa, T. & Lux, H. D. *Pflügers Arch. ges. Physiol.* **415**, 255–262 (1989).
5. Jones, O. T., Kunze, D. L. & Angelidis, K. J. *Science* **244**, 1189–1193 (1989).
6. Ozawa, S., Tsuzuki, L., Iino, M., Ogura, A. & Kudo, Y. *Brain Res.* **495**, 329–336 (1989).
7. Takahashi, K., Wakamori, M. & Akaike, N. *Neurosci. Lett.* **104**, 229–234 (1989).
8. Kasai, H., Aosaki, T. & Fukuda, J. *Neurosci. Res.* **4**, 228–235 (1987).
9. Plummer, M. R., Logothetis, D. E. & Hess, P. *Neuron* **2**, 1453–1462 (1989).
10. Tsien, R. W., Lipscombe, D., Madison, D. V., Bley, K. R. & Fox, A. P. *Trends Neurosci.* **11**, 431–438 (1988).
11. Gross, R. A., Macdonald, R. L. & Ryan-Jastrow, T. J. *Physiol., Lond.* **411**, 585–595 (1989).
12. Kamiya, H., Sawada, S. & Yamamoto, C. *Neurosci. Lett.* **91**, 84–88 (1988).
13. Krishtal, O. A., Petrov, A. V., Smirnov, S. V. & Nowycky, M. C. *Neurosci. Lett.* **102**, 197–204 (1989).
14. Fox, A. P., Nowycky, M. C. & Tsien, R. W. *J. Physiol., Lond.* **394**, 149–172 (1987).
15. Sarvey, J. M., Burgard, E. C. & Decker, G. J. *Neurosci. Meth.* **28**, 109–124 (1989).
16. Collingridge, G. L. & Bliss, T. V. P. *Trends Neurosci.* **10**, 288–293 (1988).
17. Zorumski, C. F. & Yang, J. J. *Neurosci.* **8**, 4277–4286 (1988).
18. Honoré, T., Lavridsen, J. & Kragsgaard-Larsen, P. *J. Neurochem.* **38**, 173–178 (1982).
19. Krishtal, O. A., Smirnov, S. V. & Osipchuk, Yu. V. *Neurosci. Lett.* **85**, 82–88 (1988).
20. Collingridge, G. L., Kehl, S. J. & McLennan, H. J. *Physiol., Lond.* **334**, 33–46 (1983).
21. Bertolino, M. B., Vicini, S. & Costa, E. *Neuropharmacology* **28**, 453–457 (1989).
22. Docherty, R. J. & Brown, D. A. *Neurosci. Lett.* **70**, 110–115 (1986).
23. Gähwiler, B. H. & Brown, D. A. *Neuroscience* **20**, 731–738 (1987).
24. Okada, Y. & Ozawa, S. *Eur. J. Pharmacol.* **68**, 483–492 (1980).
25. Chizmakov, I. V., Kiskin, N. I., Krishtal, O. A. & Tsyndrenko, A. Ya. *Neurosci. Lett.* **99**, 131–136 (1989).
26. Krishtal, O. A., Marchenko, S. M. & Pidoplichko, V. I. *Neurosci. Lett.* **35**, 41–45 (1983).

ACKNOWLEDGEMENTS. The work was supported by the Ukrainian Academy of Sciences and by a grant from ICI Pharmaceuticals. We are grateful to U. Heinemann for comments on the manuscript and M.C. Nowycky for help in editing the final version.

Calcium-dependent regulation of cyclic GMP phosphodiesterase by a protein from frog retinal rods

Satoru Kawamura & Motohiko Murakami

Department of Physiology, Keio University School of Medicine, Shinanomachi 35, Shinjuku-ku, Tokyo 160, Japan

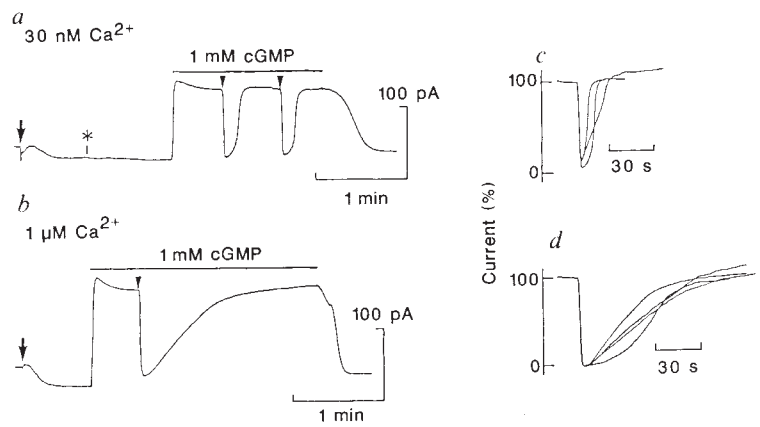
IN vertebrate photoreceptors, light reduces cyclic GMP concentration and closes cGMP-activated channels to induce a hyperpolarizing response¹. As Ca^{2+} can permeate the channels and the Na^+ - Ca^{2+} exchanger continuously extrudes Ca^{2+} , closure of the channel results in a reduction of the inter-rod Ca^{2+} concentration^{2–4}. This is believed to be one of the mechanisms of light-adaptation produced by activation of guanylate cyclase^{5–9}. Effects of Ca^{2+} on the cGMP phosphodiesterase (PDE) have been reported^{10–15}, but their physiological significance has remained unclear. We have perfused the inside-out preparation of a frog rod outer segment (I/O ROS (ref. 16), originally termed truncated ROS (ref. 17)), and find that Ca^{2+} in a physiological range regulates the light-activation of PDE. Therefore, PDE regulation by Ca^{2+} must be involved in light-adaptation in rods. The effect is mediated by a newly found protein which binds to disk membranes at high Ca^{2+} concentrations and prolongs PDE activation.

I/O ROS contains rhodopsin, transducin, PDE, guanylate cyclase and other proteins^{9,16,17}. In the present work, we applied cGMP to I/O ROS to open the cGMP-activated channels and measured PDE activity by monitoring the membrane current changes due to cGMP hydrolysis. To minimize the contribution of guanylate cyclase in the measurement of PDE activity, the GTP concentration was kept low (0.1 mM) in perfusing solutions.

Figure 1 summarizes the effect of Ca^{2+} on the time-course of the light response. I/O ROS was prepared at $1 \mu\text{M}$ Ca^{2+} and the light-responses were recorded in a solution containing cGMP at 30 nM Ca^{2+} (Fig. 1a) or $1 \mu\text{M}$ Ca^{2+} (1b). Although the light-response time-course slightly varied among specimens (Fig. 1c

FIG. 1 Evidence that Ca^{2+} is involved in the regulation of PDE light activation in rods. *a*, Membrane current record in I/O ROS at 30 nM Ca^{2+} . A frog rod was truncated at 1 μM Ca^{2+} (downward arrow), and then Ca^{2+} concentration was reduced to 30 nM (asterisk). Arrowheads, timing of the light flashes (duration 1 s). *b*, Membrane current record at 1 μM Ca^{2+} . A rod was truncated (arrow) and Ca^{2+} concentration was kept constant at 1 μM throughout. *c*, *d*, Normalized light responses recorded from three (*c*) or four (*d*) cells obtained by similar procedures as in *a* (*c*) and *b* (*d*). To avoid variation of the light-response time course between specimens, the same suction electrode was used and all the rods used in this figure were obtained from the same retina. The light stimulation was the same throughout.

METHODS I/O ROS was obtained as described previously⁹, except that the length of ROS sucked into the electrode was made short (20–25 μm) to facilitate intracellular diffusion. The suction electrode was filled with an electrode solution (115 mM choline chloride, 2.5 mM KCl, 2 mM MgCl_2 , 0.2 mM taurine, 10 mM HEPES, pH 7.5). Ca^{2+} concentration in the electrode solution was buffered at 1 μM which does not affect the I/O ROS cytoplasmic Ca^{2+} concentration appreciably⁹. ROS interior was perfused with a pseudo-intracellular solution (115 mM potassium gluconate, 2.5 mM KCl, 2 mM MgCl_2 , 0.2 mM taurine, 0.5 mM IBMX, 10 mM HEPES, 0.5–1 mM cGMP, 0.1 mM GTP, 0.2 mM ATP, pH 7.5). Under the ionic conditions used, cGMP-

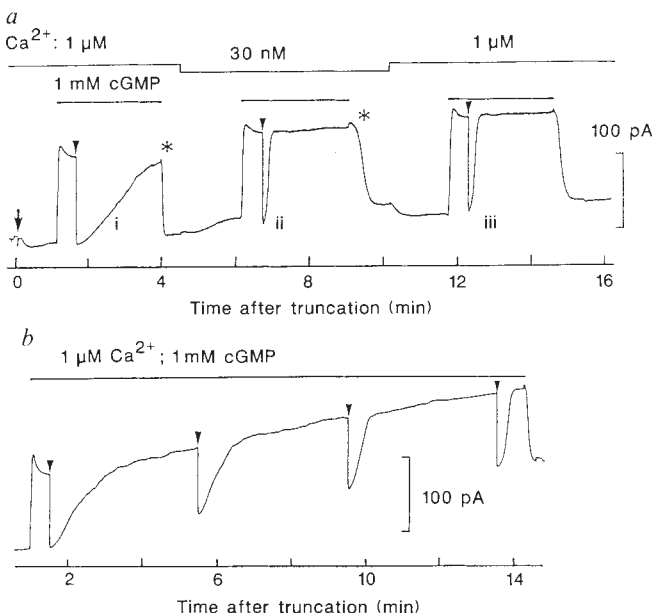


activated membrane current flows outwards. Although I/O ROS contains guanylate cyclase, cGMP synthesis in the presence of 0.1 mM GTP at 30 nM Ca^{2+} was small judging from the current amplitude in the record in Fig. 2*a*. Intracellular perfusion and Ca^{2+} buffering were performed as described previously⁹.

for 30 nM Ca^{2+} and Fig. 1*d* for 1 μM Ca^{2+}), the time course is clearly slower at 1 μM Ca^{2+} than at 30 nM Ca^{2+} .

The slow light-response at 1 μM Ca^{2+} is due to prolongation of PDE activation, because when cGMP was removed from the I/O ROS at a fixed time after a light flash, the decline of the current (the disappearance of cGMP) was much faster at 1 μM Ca^{2+} than at 30 nM Ca^{2+} (asterisks in Fig. 2*a*). This result and the slower light response can both be explained by assuming that PDE activation persists longer at 1 μM Ca^{2+} than at 30 nM Ca^{2+} . As the Ca^{2+} concentrations we used are in the physiological range^{3,4}, Ca^{2+} regulation on PDE should be one of the mechanisms of light-adaptation.

The effect of Ca^{2+} on PDE activation was irreversible: once I/O ROS was exposed to a low- Ca^{2+} solution, the light-response time course became facilitated irrespective of Ca^{2+} concentration (compare light responses *i–iii* in Fig. 2*a*). The facilitation took place even at 1 μM Ca^{2+} , although the change occurred gradually (Fig. 2*b*). Neither light, ATP nor GTP affected the facilitation. A possible mechanism for this irreversible facilitation is the elution of a soluble component which prolongs PDE activation at high Ca^{2+} . To test this possibility, we obtained a

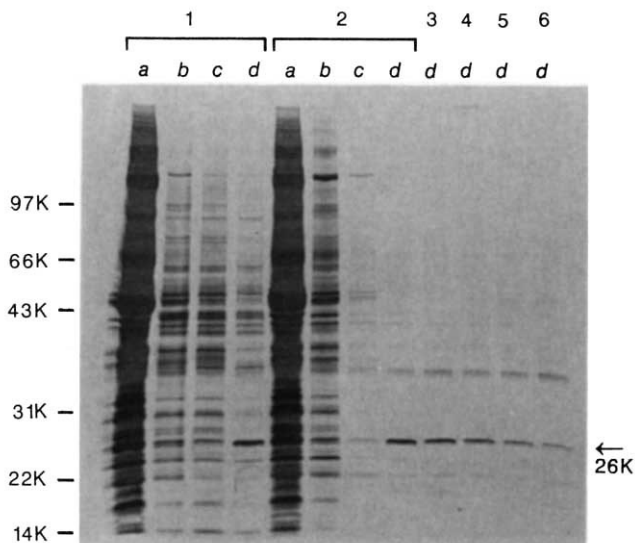


soluble crude protein fraction from a disk membrane suspension under conditions of low Ca^{2+} and added it back to I/O ROS whose light response had changed to the facilitated form by an earlier low- Ca^{2+} treatment. Addition of these proteins partially restored the Ca^{2+} sensitivity in I/O ROS (not shown), so we proceeded to purify the protein.

The result shown in Fig. 2 suggested that the protein binds to disk membranes at high Ca^{2+} but becomes soluble at low Ca^{2+} . Therefore, we first washed disk membranes with a high (1 mM) Ca^{2+} buffer solution (Fig. 3, lanes *a–c*) and then extracted proteins at low (1 nM) Ca^{2+} (low Ca^{2+} extract; Fig. 3, lanes *d*). A major band was detected with a relative molecular mass (M_r) of 26,000 (26K) (Fig. 3). The low- Ca^{2+} extracts were collected and passed over an anion exchange column (DEAE-Sephadex A-25). The 26K band was separated into two of similar M_r : first, a protein of slightly higher M_r (the 26K protein) was eluted at 10–80 mM NaCl in a 5 mM phosphate buffer and then that of lower M_r at >100 mM NaCl. The fraction containing the 26K protein showed a Ca^{2+} -dependent effect on PDE, as detected by the pH assay method⁶ (not shown).

The 26K protein was further purified by HPLC gel filtration

FIG. 2 Characterization of the Ca^{2+} effect on PDE light activation. *a*, Indications that slow recovery of the light response at 1 μM Ca^{2+} is due to prolongation of PDE activation, and that a low- Ca^{2+} treatment facilitates PDE activation time course rapidly and irreversibly. A rod was truncated (arrow) and a light flash (arrowhead) was given in the presence of 1 mM cGMP at 1 μM or 30 nM Ca^{2+} . The intensity and duration of the light flash were constant throughout the measurement. Removal of cGMP (asterisks) was carried out at a fixed time after a light flash, which ensures the precise comparison of the light effect. *b*, Gradual facilitation of PDE light activation at 1 μM Ca^{2+} . The light stimulation (arrowhead) was the same throughout the experiment.



to a single band on SDS-PAGE (Fig. 4a). The pI value is ~ 5.8 and the molar ratio to rhodopsin 1:140. The N terminus of the protein is blocked, and the partial amino-acid sequence in fragments obtained by limited proteolysis revealed some similarity to visinin¹⁸ and calbindin^{19,20}. Light had no effect on the Ca^{2+} -dependent binding of the 26K protein to disk membranes.

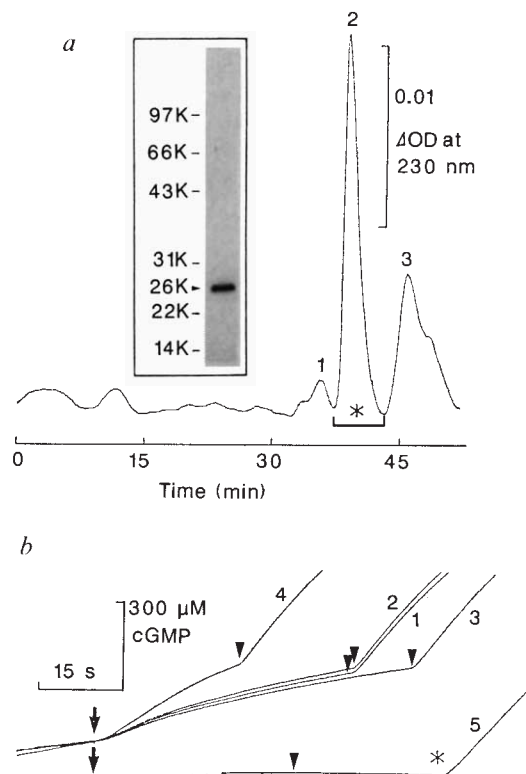
The purified 26K protein was reconstituted with washed disk membranes, and PDE activation was monitored by the pH assay method⁶ (Fig. 4b). Without the 26K protein, PDE activation was not influenced significantly by Ca^{2+} (record 1 for 30 nM Ca^{2+} and record 2 for 1 μM Ca^{2+}). Addition of the protein ($\sim 2 \mu\text{M}$) did not affect PDE activation significantly at 30 nM Ca^{2+} (record 3). However, the 26K protein increased the efficiency of PDE activation by more than 50% at 1 μM Ca^{2+} without affecting the maximum PDE activity (record 4). We eliminated the contribution of guanylate cyclase for this

FIG. 3 Extraction of proteins which bind to disk membranes at high Ca^{2+} . Frog retinas were isolated with the aid of an infra red-converter. ROSs were brushed off retinæ in a 1 mM Ca^{2+} buffer solution (115 mM potassium gluconate, 2.5 mM KCl, 2 mM MgCl_2 , 1 mM CaCl_2 , 1 mM DTT, 10 mM HEPES, pH 7.5). Most of the ROSs were disrupted during this process. The soluble protein fraction (high- Ca^{2+} extract) was separated from membranes by centrifugation at 200,000g for 15 min. An aliquot of the high- Ca^{2+} extract was subjected to SDS-PAGE (lane 1a). Disk membranes were purified with 43% (w/w) sucrose floatation, and then washed twice by centrifugation with the 1 mM Ca^{2+} buffer solution. SDS-PAGE patterns of the supernatant fractions are shown in lanes 1b and c. Proteins which bind to disk membranes at high Ca^{2+} were extracted with a 1 mM Ca^{2+} buffer solution (115 mM potassium gluconate, 2.5 mM KCl, 2 mM MgCl_2 , 0.1 mM CaCl_2 , 2.78 mM EGTA, 1 mM DTT, 10 mM HEPES, pH 7.5) (lane 1d). As extractable proteins still remained in the high- Ca^{2+} extract, we repeated low- Ca^{2+} extraction by adding the high- Ca^{2+} extract to the disk membranes (lanes 2a–6d). After the third cycle, only the low- Ca^{2+} extracts (d) were applied to the gel. Proteins were silver-stained. The major components of the low- Ca^{2+} extract were two 26K proteins which could be separated by a DEAE column (see text). Both proteins reversibly bound to disk membranes when the Ca^{2+} concentration increased.

measurement because without cGMP we did not detect hydrolysis of cGMP (record 5). Considering the 26K protein/rhodopsin ratio and the estimated extrarod volume of ROS, the 26K protein concentration in intact rods is nearly 40 μM . The protein therefore has a much greater effect in intact cells than in the experiment in Fig. 4b. Although the low yield of the 26K protein (15–20%; Fig. 4a legend) hampered the measurement of the dose-response relation at the physiological range of the protein, we observed an increase in the efficiency of PDE activation up to 9.2 μM using a partially purified preparation.

In the dark-adapted state, the inter-rod calcium concentration ($[\text{Ca}^{2+}]_i$) is relatively high and thus the 26K protein is expected to bind to the disk membrane to prolong PDE activation. During light-adaptation, $[\text{Ca}^{2+}]_i$ becomes low and the protein is released from the membrane to cause a reduction in the efficiency of PDE activation. Therefore, the 26K protein regulates the rod

FIG. 4 a, Purification of the 26K protein and b, the reconstitution experiment. METHODS. a, Proteins eluted at 10–80 mM NaCl from a DEAE column (see text) were collected and subjected to HPLC gel filtration (TSKgel G2000SW + G3000SW; Tosoh). Protein elution was monitored by absorbance at 230 nm. Peak 2 corresponded to elution of the 26K protein and the fractions labelled with an asterisk were collected. SDS-PAGE pattern with silver staining ($\sim 0.3 \mu\text{g}$ protein) is shown in the inset. In the condensation process during purification, we used ultra filters (UFP2 LGC, Millipore). These filters irreversibly adsorbed the 26K protein, which reduced its yield. In the later work, therefore we treated the filter with ovalbumin before use to reduce the adsorption of the 26K protein. During condensation, we also added ovalbumin (200 μg) which did not affect PDE light activation (see also ref. 21). The overall yield of the 26K protein was 15–20%, assuming that the molar ratio to rhodopsin is 1:140 (see text). b, Frog disk membranes were obtained in the dark and washed with the 1 mM Ca^{2+} buffer to eliminate all of the soluble proteins including the 26K protein. The membranes were kept in a refrigerator overnight after addition of 0.2 mM ATP. Although washing of frog disk membranes increases dark PDE activity²², the above procedure successfully reduced the dark activity. GTP and ATP (final concentrations of both, 0.45 mM), and cGMP (final concentration, 3.6 mM) were added to the washed membranes (before the records shown in the figure). A weak light-flash (arrow) bleaching 3.5×10^{-5} fraction of rhodopsin was first given, and then a continuous intense light (arrowhead) bleaching 1.4% of rhodopsin per s was given. The pH drop induced by hydrolysis of cGMP was monitored. Assay volume, 200 μl (ref. 11); rhodopsin concentration, 10 μM . Measurements were performed in the absence of the 26K protein (record 1 for 30 nM Ca^{2+} and record 2 for 1 μM Ca^{2+}), and in its presence ($\sim 2 \mu\text{M}$) (record 3 for 30 nM Ca^{2+} and record 4 for 1 μM Ca^{2+}). As a control, cGMP was added only after light stimulation (record 5, asterisk). The fraction of PDE light activation caused by the weak light flash was 30.1% (record 1), 31.8% (2), 23.8% (3) and 46.0% (4).



light sensitivity in a Ca^{2+} -dependent manner. We propose to call this protein 'sensitivity-modulating protein (S-modulin)'. Because guanylate cyclase activity is also a function of Ca^{2+} (refs 5-9), a single mechanism, $[\text{Ca}^{2+}]_i$ decrease, regulates both synthesis and hydrolysis of cGMP in rods. □

Received 11 September; accepted 21 November 1990.

1. Fesenko, E. E., Kolesnikov, S. S. & Lyubarsky, A. L. *Nature* **313**, 310-313 (1985).
2. Yau, K.-W. & Nakatani, K. *Nature* **313**, 579-582 (1985).
3. McNaughton, P. A., Cervetto, L. & Nunn, B. J. *Nature* **322**, 261-263 (1986).
4. Ratto, G. M., Payne, R., Owen, W. G. & Tsien, R. Y. *J. Neurosci.* **8**, 3240-3246 (1988).
5. Hodgkin, A. L. & Nunn, B. J. *J. Physiol., Lond.* **403**, 439-471 (1988).
6. Koch, K. & Stryer, L. *Nature* **334**, 64-66 (1988).
7. Matthews, H. R., Murphy, R. L. W., Fain, G. L. & Lamb, T. D. *Nature* **334**, 67-69 (1989).
8. Nakatani, K. & Yau, K.-W. *Nature* **334**, 69-71 (1989).
9. Kawamura, S. & Murakami, M. *J. gen. Physiol.* **94**, 649-668 (1989).
10. Robinson, P. R., Kawamura, S., Abramson, B. & Bownds, M. D. *J. gen. Physiol.* **76**, 631-645 (1980).
11. Kawamura, S. & Bownds, M. D. *J. gen. Physiol.* **77**, 571-591 (1981).
12. Del Priore, L. V. & Lewis, A. *Biochem. biophys. Res. Commun.* **113**, 317-324 (1983).
13. Wagner, R., Ryba, N. & Uhl, R. *FEBS Lett.* **242**, 249-254 (1989).
14. Barkdoll, A. E., Pugh, E. N. Jr & Sitarumayya, A. *J. gen. Physiol.* **93**, 1091-1108 (1989).
15. Torre, V., Matthews, H. R. & Lamb, T. D. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7109-7113 (1986).
16. Kawamura, S. & Murakami, M. *Neurosci. Res. Suppl.* **10**, S15-S22 (1989).
17. Yau, K.-W. & Nakatani, K. *Nature* **317**, 252-255 (1985).
18. Yamagata, K., Goto, K., Kuo, C.-H., Kondo, H. & Miki, N. *Neuron* **2**, 469-476 (1990).
19. Wilson, P. W., Harding, M. & Lawson, D. E. M. *Nucleic Acids Res.* **13**, 8867-8881 (1985).
20. Yamakuni, T. et al. *Nucleic Acids Res.* **14**, 6768 (1986).
21. Buzdygon, B. E. & Liebman, P. A. *J. biol. Chem.* **259**, 14567-14571 (1984).
22. Kawamura, S. & Murakami, M. *Biochim. biophys. Acta* **870**, 256-266 (1986).

ACKNOWLEDGEMENTS. We thank H. Asou, K. Takamatsu and K. Uemura for help in the biochemical experiments. Supported by the Ministry of Education, Science and Culture of Japan (Priority Areas).

Cloning and characterization of the cDNAs for human and rat corticotropin releasing factor-binding proteins

E. Potter^{*†}, D. P. Behan^{*‡}, W. H. Fischer^{*},
E. A. Linton[‡], P. J. Lowry[‡] & W. W. Vale^{*}

^{*} Clayton Foundation for Peptide Biology, Salk Institute, La Jolla, California 92138, USA

[†] Department of Pharmacology, University of California at San Diego, La Jolla, California 92037, USA

[‡] Department of Biochemistry and Physiology, School of Animal and Microbial Sciences, Reading University, Reading RG6 2AJ, UK

CORTICOTROPIN-releasing factor (CRF)¹, is a potent stimulator of synthesis and secretion of prepromelanocortin-derived peptides. Although CRF concentrations in the human peripheral circulation are normally low²⁻⁴, they increase throughout pregnancy⁴⁻⁸ and fall rapidly after parturition. Maternal plasma CRF probably originates from the placenta, which responds to the bioactive peptide^{5,9,10} and produces the peptide⁹ and its messenger RNA¹¹. Even though CRF concentrations in late gestational maternal plasma are similar to those in rat hypothalamic portal blood^{12,13} and to those that can stimulate release of adrenocorticotrophic hormone (ACTH) *in vitro*, maternal plasma ACTH concentrations increase only slightly with advancing gestation and remain within the normal range¹⁴. Several groups have now reported the existence of a CRF-binding protein in human plasma which inactivates CRF¹⁵⁻²⁰ and which has been proposed to prevent inappropriate pituitary-adrenal stimulation in pregnancy. The binding protein was recently purified from human plasma¹⁹. We have now isolated and partially sequenced the binding protein, allowing us to clone and characterize its complementary DNA from human liver and rat brain. Expression of the cDNAs for human and rat binding protein in COS7 cells showed that these proteins bind CRF with the same affinity as the native human protein¹⁵. Both rat and human recombinant binding proteins inhibit CRF binding to a CRF antibody and inhibit CRF-induced ACTH release by pituitary cells *in vitro*.

Human CRF-binding protein (hCRF-BP) was purified as previously described¹⁹ and subjected to SDS gel electrophoresis. The 37K band (relative molecular mass (M_r) 37,000) was trypsinized and the resultant fragments separated on reversed-phase HPLC and sequenced by Edman degradation. Sequence information was obtained from seven of these fragments (Fig. 1), and those with the lowest degeneracy were chosen for synthesis of oligonucleotide probes. DNA from an adult human liver cDNA library was used as template for the polymerase chain reaction (PCR) using various combinations of the sense and antisense sequences from the oligonucleotides generated from tryptic fragments 1 and 5. The PCR reaction generated a DNA fragment of 387 base pairs (bp), which when subcloned and sequenced contained an open reading frame encoding tryptic fragments 2, 3 and 4, along with peptides 1 and 5.

A full-length cDNA containing a 1.8-kilobase (kb) insert was obtained by screening a second adult liver cDNA library. The predicted open reading frame codes for a protein of 322 amino acids (Fig. 1), containing all the amino-acid sequences from the tryptic fragments of the purified hCRF-BP and unrelated to any previously known protein. The finding of a putative *N*-glycosylation site (at amino acid position 203) is in agreement with the presence of asparagine-linked sugar moieties in native hCRF-BP²¹. Analysis of the full-length sequence revealed no evidence for long hydrophobic stretches which could represent transmembrane regions of a putative membrane receptor²². There are 10 interdispersed cysteine residues (excluding the signal sequence) suggesting the presence of five intramolecular disulphide bonds.

After detecting binding protein mRNA in rat brain we screened a rat cortex cDNA library and isolated several clones. The sequence of one clone containing a 1.85-kb insert predicts a protein of 322 amino acids which is 84% identical to the hCRF-BP. The cysteine residues and the putative *N*-glycosylation site are conserved between the rat and human sequences.

Northern blot analysis of human adult and fetal liver mRNA identified a 1.9-kb transcript for hCRF-BP (Fig. 2). A similarly sized transcript for CRF-BP is present in human placenta, fetal monkey brain and adult rat cortex.

Okayama-Berg vectors were prepared for expression studies on transfected COS7 cells (Fig. 3 legend). Cells that were transfected with the cDNAs for either hCRF-BP or rat (r) CRF-BP, but not those transfected with vector only, secreted a protein into the medium which could inhibit CRF binding to an anti-CRF antibody (Fig. 3). In addition, the conditioned media also contained activities which inhibited CRF-induced ACTH release from primary rat pituitary cells in a competitive manner (Fig. 4) presumably as a consequence of binding to CRF. However, CRF has effects on many biological systems and we cannot rule out the possibility that the peptide might retain some non-hypophyseal actions even when associated with the binding protein.

Scatchard analyses (data not shown) of the human and rat CRF-BPs show that the recombinant binding proteins have the same affinity for human or rat CRF (which are identical to one another) as does the purified human CRF-BP ($K_d = 0.1 \pm 0.2$ nM). The recombinant human and rat CRF-BPs bind ovine CRF with a much lower affinity (250 nM) suggesting that one or more of the seven amino acids that differ between the human/rat and ovine forms of CRF are involved in the binding to the CRF-BP. This result indicates a major difference in the structural requirements for binding to CRF-BP and to the pituitary CRF receptor, which does not distinguish between human/rat CRF and ovine CRF. The CRF-BPs have as high or higher affinity for human/rat CRF than do CRF receptor preparations which have dissociation constants (K_d) of between 0.1 and 0.5 nM²³.

Although human and rat CRF-BP are highly homologous, they show distinct patterns of expression in their respective organisms. The primate CRF-BP gene is expressed in liver,

- a**
- 1 G355 EAADYD(P)FLLF(S)ANLK
 - 2 G353 ELAGEQPYR
 - 3 G351 VFDGWILK
 - 4 G372 FPSSQDHPLPSAEXR
 - 5 G352 (S)(C)QNV A MIFF(R)
 - 6 G357 VHE(P)GNGF(T)L(T)IK
 - 7 G354 VTFEYR

b

Hum	AAAGAGACCCAGGAAAGGACCTAGCAGCTTCGAGTTCTCAGTGTGGGCGAAGGCGAGGGAAGAAACGCCTAA	GGACCTCCGGAGCAGACAGCAGCAGCTGCAGAGGCAAGGCCAG•C	-1
Rat		T C C •• T C CA	-1
Hum	ATG TCG CCC AAC TTC AAA CTT CAG TGT CAC TTC ATT CTC ATC TTC CTG ACG GCT CTA AGA GGG GAA AGC CGG TAC CTA GAG CTG AGG GAA		90
Rat	M S P N F K L C A C H F I L I C L T A L R G A G S R	Y L E G CAA	30
Rat		V Q	
Hum	GCG GCG GAC TAC GAT CCT TTC CTG CTC TTC AGC GCC AAC CTG AAG CGG GAC GTG GCT GGG GAG CAG CCG TAC CGC CGC GCT CTG CGG TGC		180
Rat	A A A D Y D P F L L F S A N T L K R	D V A G E Q P Y R A R A L R C	60
Rat		N L E	
Hum	CTG GAC ATG CTG AGC CTC CAG GGC CAG TTC ACC TTC ACC GCC GAC CGG CCG CAG CTG CAC TGC GCA GCC TTC TTC ATC AGC GAG CCC GAG		270
Rat	L D M L S L Q G Q F T F T A D R P Q L H C A A F F I S E P E		90
Rat			
Hum	GAG TTC ATT ACC ATC CAC TAC GAC CAG GTC TCC ATC GAC TGT CAG GGC GGC GAC TTC CTG AAG GTA TTT GAT GGT TGG ATT CTC AAG GGG		360
Rat	E F I C T I H Y TT D Q V S I D C Q G T G D F L K	V F D C T W I L K G	120
Rat			
Hum	GAG AAG TTC CCC AGT TCC CAG GAT CAT CCT CTC CCC TCA GCT GAG CGG TAC ATA GAT TTC TGT GAG AGT GGT CTT AGC AGG AGG AGC ATC		450
Rat	E K F P S S Q D H P L P A C AGG A A Y I D F C E S G L S R R S I V		150
Rat			
Hum	AGA TCT TCC CAG AAT GTG GCC ATG ATC TTC TTC CGA GTC CAT GAA CCA GGA AAT GGA TTC ACA TTA ACC ATA AAG ACA GAC CCC AAC CTC		540
Rat	R S S Q N V A M I F F R V H E P G N G F T L T I K T D P N L		180
Rat			
Hum	TTT CCT TGC AAT GTC ATT TCT CAG ACT CCA AAT GGA AAG TTT ACC CTG GTA GTT CCA CAC CAG CAT CGA AAC TGC AGC TTC TCC ATA ATT		630
Rat	F P C N V I S Q T P N G K F T L V V P H Q H C S F S I I		210
Rat			
Hum	TAT CCT GTG GTG ATC AAA ATA TCT GAT CTT ACC CTG GGA CAC GTA AAT GGT CTT CAG TTA AAG AAA TCC TCA GCA GGT TGC GAG GGA ATA		720
Rat	Y P V V I K I S D L T A L G H V N G L Q L K K C T G G T C T G T G T G T		240
Rat			
Hum	GGA GAC TTT GTG GAG CTG CTG GAG GGA ACT GGA TTG GAC CCT TCC AAG ATG ACG CCT TTA GCT GAT CTC TGC TAC CCC TTT CAT GGC CCG		810
Rat	G D F V E L L E G T G L D S K M T T C A D L C Y P F H G T		270
Rat			
Hum	GCC CAG ATG AAA GTT GGC TGT GAC AAC ACT GTG GTG CGC ATG GTC TCC AGT GGA AAA CAC GTA AAT CGT GTG ACT TTT GAG TAT CGT CAG		900
Rat	A Q M K V G C D N T V V R M V S S G K H V N R V T F E Y R Q		300
Rat			
Hum	CTG GAG CCG TAC GAG CTG GAA AAC CCA AAT GGA AAC AGT ATC GGG GAA TTC TGT TTG TCT GGT CTTGAATAACCAACCCAGTGATTACATGCTGAT		998
Rat	L E P Y E L E N P N G N S I G E F C L S S		322
Rat			
Hum	AGCTAAGTGAGTTTAAATGGCCATTGTGTATGATTTTGTATGCACAAGTAAAGCCTTTCATACCACTCAGTATTTCCCGCCTTGAGCGCAGCAGACACACACACATACACA		1117
Hum	CACGCATTATTTTGTACTTTGCTTTTATGTTTGTATCTGTAAATGAACACATGGCAGAAAATAACCTGATTGGTAGG		1202

FIG. 2 **a**, Northern blot analysis of human CRF-BP. Lane 1, RNA ladder; 2, total fetal liver (20 μ g). **b**, Northern blot analysis of rat tissues with rat CRF-BP. Lane 1, adrenals; 2, brain; 3, heart; 4, intestines; 5, kidney; 6, liver; 7, lung; 8, muscle; 9, spleen; 10, testis; 11, no RNA; 12, RNA ladder. Each lane contains 5 μ g poly(A)⁺ RNA.

METHODS. RNA was electrophoresed on 1.5% agarose gels containing 6% formaldehyde³⁰. Protocols for blotting, prehybridization and hybridization are as described³⁰. Inserts were ³²P-labelled with random primers and Klenow (USB). The probe for panel **a** was the *Eco*RI fragment from the partial human cDNA clone pHL9A and for panel **b**, was the *Pst*II fragment from rat cDNA clone pRC2A.

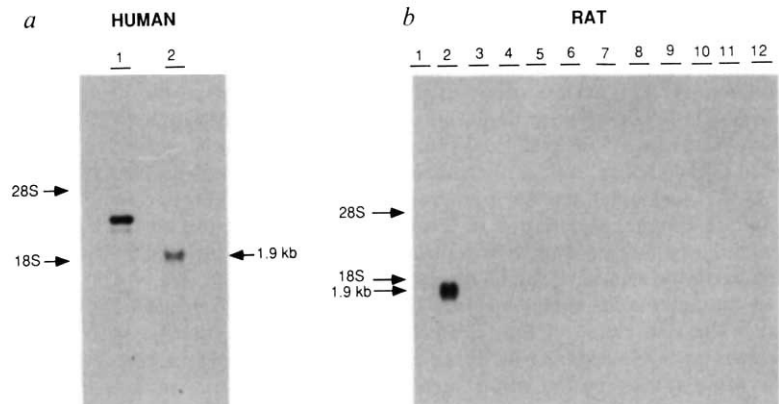


FIG. 1 a, Amino-acid sequences determined for seven tryptic fragments. b, Complete cDNA and predicted amino-acid sequence for human and rat CRF-BP. Numbers below each line are differences of the rat sequence from human. The nucleotide sequence is numbered in the 5' to 3' direction beginning with the first ATG of the open reading frame. The first ATG triplet downstream of nonsense codons found in-frame with the rest of the coding block and having a Kozak consensus sequence²⁴ was assigned as the site of translation. Numbering is at the right hand of each line. The underlined amino-acid sequences are those determined from tryptic fragments. N-terminal sequence analysis of the intact protein revealed the presence of two species: (Y, E) (L, A) (E, A) (L, D) Y (E, D) (A, P) (A, F) (single-letter amino-acid code). The occurrence of a mixed sequence was attributed to partial proteolytic processing at Arg 29 and thus the N-terminal amino acid was deduced to be Tyr 25 which is represented by the box. There is one possible N-glycosylation site at amino acid 204, identified by a circle. M_r of CRF-BP is 37K and the deduced M_r of the cloned protein is 30K which agrees with that of the unglycosylated form of the protein. We have found no known proteins homologous to human CRF-BP.

METHODS. a, Protein for sequencing was purified as described¹⁹. About 100 pmol was further purified by SDS gel electrophoresis and transferred to nitrocellulose. The protein band corresponding to CRF-BP was excised and treated with trypsin *in situ* as described^{25,26}. Tryptic fragments were resolved by reversed-phase HPLC and each Edman-degraded²⁶. N-terminal sequence analysis was carried out on protein purified by SDS gel electrophoresis and electrotransferred to Immobilon²⁷. b, Cloning of human CRF-BP. Sequences of the oligonucleotide primers corresponding to tryptic fragments 1 and 5 were GA(T/C)TA(T/C)GATCCTTT(T/C)TTNGTNTT(T/C)(T/A)(C/G)NGCNAAC and CA(A/G)AA(T/C)GTGNCNATGATNTT(C/T)TTC. DNA from a human adult liver cDNA library²⁸ served as template in 25 cycles of PCR with 1 min denaturation at 94 °C, 2 min of annealing at 45 °C, and 3 min of extension at 72 °C. The PCR products were analysed on a 1% (w/v) TBE agarose gel; a 387-bp fragment was electroeluted into 12 M ammonium acetate solution using a IBI model UEA Bio-Rad electroeluter. The 387 PCR fragment was subcloned into the *Sma*I site of Bluescript KS vector (pHL9). Nucleotide sequencing was by the Sanger dideoxy chain termination method using Sequenase (USB). Random primed pHL9 insert was used to screen the original human liver cDNA library. Duplicate nitrocellulose filters were hybridized in 50% formamide, 5 × SSC buffer, 1 × Denhardt's solution, 0.1% SDS, 100 µg ml⁻¹ sheared salmon sperm DNA and ³²P-labelled insert (10⁶ c.p.m. ml⁻¹) 42 °C for 18 h. Filters were washed at 60 °C in 2 × SSC. Two partially overlapping clones for the hCRF-BP coding region were isolated (pHL9A and pHL7J) containing inserts of 650 and 570 bp respectively. A new library using adult human liver RNA was constructed to obtain full-length cDNA clones. Messenger RNA was isolated by the guanidium isothiocyanate/caesium chloride method²⁹ and oligo (dT) chromatography. mRNA (10 µg) was used for the λZap II cloning system (Stratagene) which made a library of 5 × 10⁶ plaque-forming units in total. Plaques (10⁶) were screened with inserts from pHL9A and pHL7J as described above. Clone pHLA3 was isolated and the sequence is presented in the figure. The rat gene, using the insert from clone pHL9A as a probe, was isolated from the Okayama-Berg rat cortex library (gift of S. Gould) under the following conditions: 10⁶ colonies were plated out on ampicillin plates, nitrocellulose filters were lifted after 8 h and amplified on chloramphenicol plates for 18 h. Filters were hybridized as described above. Six positives were isolated. Clone pRC2A contained a 1.8-kb insert whose sequence is presented in the figure.

placenta and brain whereas in the rat we have observed binding protein mRNA only in the brain. This difference in tissue expression could reflect an adaptation in primates which prevents inappropriate pituitary effects of placental CRF during gestation. In contrast, the rodent placenta does not produce significant amounts of CRF and, therefore, a binding protein would not be required to protect the pituitary.

CRF or its target-cell receptors are broadly distributed throughout the central nervous system and in several peripheral tissues including placenta, adrenal, sympathetic ganglia, lymphocytes, gastrointestinal tract, pancreas and gonads²³. Generally, CRF is produced and acts locally in a transynaptic, paracrine or neuroendocrine fashion. The finding of CRF-BP mRNA in the brains of primates and rats raises the possibility

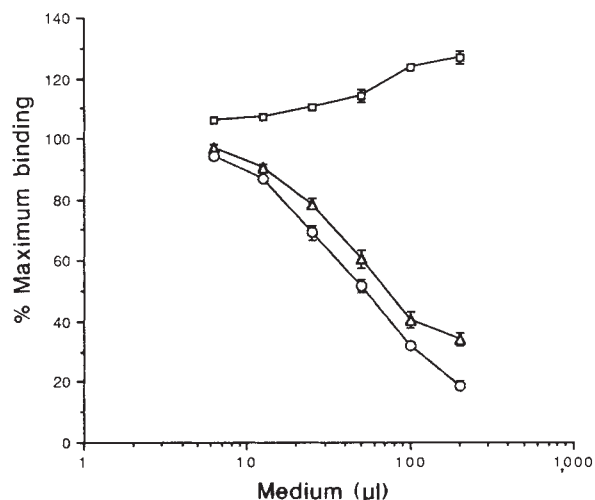


FIG. 3 Inhibition of CRF binding to anti-CRF serum by recombinant CRF-BP. Data are presented as percentage of maximum binding. □, Control media of cells transfected with pSG5 vector with no insert; ○, media from cells transfected with rat CRF-BP (pRC2A); △, media from cells transfected with human CRF-BP (pSG5-hA3).

METHODS. The full length rCRF-BP clone was isolated from an Okayama-Berg library and thus contained a simian virus 40 (SV40) promoter and polyadenylation signal and is suitable for transfection into COS7 cells³¹. The human CRF-BP expression vector was constructed by digesting the plasmid pHLA3 with *Xho*I, repairing with Klenow and releasing the hCRF-BP coding region with *Bam*HI. The insert was cloned into the SV40 expression vector pSG5 (Stratagene) which had been digested with *Bgl*II, and repaired with Klenow and *Bam*HI. COS7 cells were transiently transfected using DEAE-dextran³² with SV40 expression vectors containing cDNA inserts for human and rat CRF-BP or with vector only, and media was collected after 72 h. Various dilutions of conditioned media were incubated with [¹²⁵I]CRF trace and a 1:6,000 dilution of rabbit anti-rCRF antibody (rC67) for 30 min at room temperature. Samples were precipitated with sheep anti-rabbit γ-globulin and 10% polyethylene glycol, washed with SPEA buffer (50 mM sodium phosphate, 100 mM sodium chloride, 25 mM EDTA, 0.1% sodium azide), centrifuged and the pellets counted.

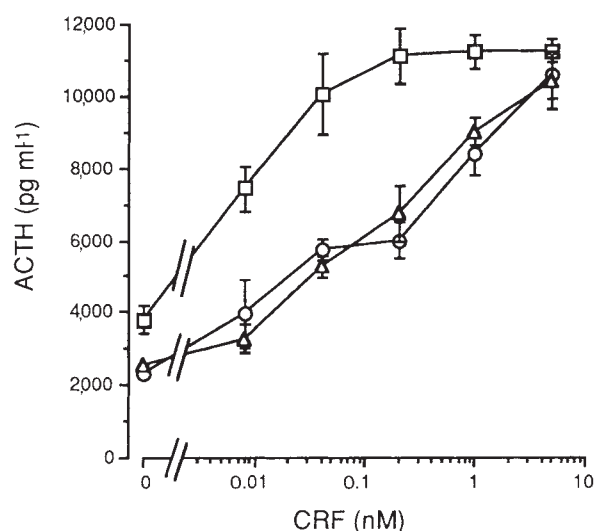


FIG. 4 Inhibition of CRF-induced ACTH release from anterior pituitary cell cultures. □, Control media from cells transfected with pSG5 vector with no insert; ○, media from cells transfected with rat CRF-BP; △, media from cells transfected with human CRF-BP.

METHODS. COS7 cells were transiently transfected using DEAE-dextran with SV40 expression vectors containing cDNA inserts for human and rat CRF-BP or with vector only, and media was collected after 72 h. Conditioned media was placed on primary anterior pituitary cell cultures as previously described³³, varying concentrations of rCRF were added to the media and cultures were incubated for 3 h. Media was removed and assayed for ACTH by double-antibody radioimmunoassay (Diagnostic Products).

that this protein might colocalize to some CRF pathways and modulate the neural actions of this neuropeptide. □

Received 9 October; accepted 30 November 1990.

- Vale, W., Spiess, J., Rivier, C. & Rivier, J. *Science* **213**, 1394-1397 (1981).
- Cunnah, D., Jessop, D., Besser, G. M. & Rees, L. H. *J. Endocr.* **113**, 123-131 (1987).
- Linton, E. A. et al. *J. clin. Endocr. Metab.* **64**, 1047-1053 (1987).
- Suda, T., et al. *J. clin. Invest.* **76**, 2026-2029 (1985).
- Sasaki, A., et al. *J. clin. Endocr. Metab.* **59**, 812-814 (1984).
- Goland, R. S., Wardlaw, S. L., Stark, R. I. Jr, Brown, L. S. B. & Frantz, A. G. *J. clin. Endocr. Metab.* **63**, 1199-1203 (1986).
- Laatikainen, T., Virtanen, T., Raisanen, I. & Salminen, K. *Neuropeptides* **10**, 343 (1987).
- Campbell, E. A., et al. *J. clin. Endocr. Metab.* **64**, 1054-1059 (1987).
- Petragnia, F., Sawchenko, P., Rivier, J. & Vale, W. *Nature* **328**, 717 (1987).
- Shibasaki, T., Odagiri, E., Shizume, K. & Ling, N. *J. clin. Endocr. Metab.* **55**, 384-386 (1982).
- Grino, M., Chrousos, G. P. & Margioris, A. S. *Biochem. biophys. Res. Commun.* **148**, 1208-1214 (1987).
- Plotsky, P. M. & Sawchenko, P. E. *Endocrinology* **120**, 1361-1369 (1987).
- Fink, G., Robinson, K. A. F. & Tannahill, L. A. *J. Physiol., Lond.* **401**, 329 (1988).
- Rees, L. H., Burke, C. W., Chard, T., Evans, S. W. & Letchworth, A. T. *Nature* **254**, 620-622 (1975).
- Orth, D. N. & Mount, C. D. *Biochem. biophys. Res. Commun.* **143**, 411-417 (1987).
- Suda, T., et al. *J. clin. Endocr. Metab.* **67**, 1278 (1988).
- Linton, T. A., Wolfe, C. D. A., Behan, D. P. & Lowry, P. J. *J. Clin. Endocr.* **28**, 315-324 (1988).
- Ellis, M. J., Livesey, J. H. & Donald, R. A. *J. Physiol., Lond.* **117**, 299 (1988).
- Behan, D. P., Linton, E. A. & Lowry, P. J. *J. Endocr.* **122**, 23-31 (1989).

- Linton, E. A., Behan, D. P., Saphier, P. W. & Lowry, P. J. *J. clin. Endocr. Metab.* **70**, 1574-1580 (1989).
- Suda, T., Sumitomo, T., Tozawa, F., Ushiyama, T. & Demura, H. *Biochem. Biophys. Res. Commun.* **165**, 703-707 (1989).
- Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105 (1982).
- DeSouza, E. B. & Nemeroff, C. B. *Corticotropin-releasing Factor: Basic and Clinical Studies of a Neuropeptide* (CRC Press, Boca Raton, FL, 1990).
- Kozak, M. *Cell* **44**, 283 (1986).
- Aebersold, R. H., Leavitt, J., Saavedra, R. R., Hood, L. E. & Kent, S. B. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6970 (1987).
- Fischer, W. H., Karr, D., Jackson, B., Park, M. & Vale, W. *Meth. Neurosci.* (in the press).
- Matsudaira, P. *J. biol. Chem.* **262**, 10035 (1987).
- Kwok, S. C. M., Ledley, F. D., DiLella, A. G., Robson, K. J. H. & Woo, S. L. C. *Biochemistry* **24**, 556-561 (1985).
- Chirgwin, J., Przbyla, A., MacDonald, R. & Rutter, W. *Biochemistry* **18**, 5294-5299 (1979).
- Ausubel, F. et al. *Current Protocols in Molecular Biology* (Greene/Wiley-Interscience, New York, 1989).
- Okayama, H. & Berg, P. *Molec. cell. Biol.* **2**, 161-170 (1982).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).
- Vale, W. et al. in *Meth. Enzymol.* **124**, 389-401 (1986).

ACKNOWLEDGEMENTS. This research was supported by the NIH, the Adler Foundation, the Robert J. Kieberg, Jr and Helen C. Kieberg Foundation, the Clayton Foundation for Research, California Division, the MRC and the Reading University Endowment Fund. W.V. and W.F. are Clayton Foundation Investigators. We thank A. Kakizuka for the generous gift of the northern blot of rat tissues, L. Mathews and R. Evans for their critical reading of the manuscript, and C. Douglas, C. Donaldson and B. Jackson for technical assistance.

Raf-1 protein kinase is required for growth of induced NIH/3T3 cells

Walter Kolch*, Gisela Heidecker*, Patricia Lloyd† & Ulf R. Rapp*‡

*Laboratory of Viral Carcinogenesis, NIH/NCI, Frederick Cancer Research and Development Center, Frederick, Maryland 21702-1201, USA
†Biological Carcinogenesis and Development Program, Program Resources, Inc./DynCorp, NCI, Frederick Cancer Research and Development Center, Frederick, Maryland 21702-1201, USA

MANY growth factors regulate the cytoplasmic Raf-1 protein kinase¹⁻¹⁰, consistent with its having a central role in transduction of growth signals. The kinase is ubiquitously expressed¹¹ and can promote proliferation¹², presumably in a manner dependent on growth-factor receptors and membrane-associated oncogenes¹³⁻¹⁵. We have now examined the dependence of serum- and TPA (12-*O*-tetradecanoylphorbol-13-acetate)-regulated NIH/3T3 cell growth on RAF-1 kinase to determine whether Raf-1 is essential for receptor signalling. We inhibited Raf-1 function by expressing *c-raf-1* antisense RNA or kinase-defective *c-raf-1* mutants. Antisense RNA for *c-raf-1* interferes with proliferation of normal NIH/3T3 cells and reverts *raf*-transformed cells. In revertant cells, DNA replication induced by serum or TPA was eliminated or reduced proportionately to the reduction in Raf protein levels. Expression of a kinase-defective Raf-1 mutant (craf301) or a regulatory domain fragment (HCR) inhibited serum-induced NIH/3T3-cell proliferation and *raf* transformation even more efficiently. Inhibition by antisense RNA or craf301 blocked proliferation and transformation by Ki- and Ha-ras oncogenes. We conclude that *raf* functions as an essential signal transducer downstream of serum growth factor receptors, protein kinase C and *ras*.

Portions of *c-raf-1* cDNAs were expressed in sense and antisense orientation using the pMNC vector (Fig. 1). After transfection into NIH/3T3 cells the number of neomycin-resistant colonies was scored. Antisense constructs yielded roughly half the number of colonies as did the corresponding sense construct or the vector control, indicating that *raf* antisense RNA interferes with viability and/or proliferation. As NIH/3T3 cells express no B-raf and 10-fold less A-raf than Raf-1 (ref. 11), we ascribe the effect to interference with Raf-1. Antisense colonies

TABLE 1 *raf*-inhibition blocks *ras*-mediated proliferation and transformation

a) v-Ki-ras cell transfection

Plasmids	Yield of neo ^r		Morphology of neo ^r colonies	
	colonies	flat	intermediate	transformed
pMNC*	100 ± 0%	0 ± 0%	0 ± 1%	100 ± 1%
pMNC301-2 [†]	61 ± 8%	2 ± 1%	15 ± 7%	83 ± 7%
pMNC301-1 [‡]	30 ± 7%	15 ± 3%	15 ± 5%	70 ± 6%

b) NIH/3T3 co-transfection with v-Ha-ras (pSV2neo/ras) and p301

Plasmids (ratio 1:1)	Inhibition [§]	Morphology of neo ^r colonies		
		flat	intermediate	transformed
<i>ras</i> + pMNC*	0 ± 3%	27 ± 3%	17 ± 5%	56 ± 3%
<i>ras</i> + p301-2 [†]	53 ± 4%	28 ± 1%	46 ± 9%	26 ± 7%
<i>ras</i> + p301-1 [‡]	61 ± 3%	46 ± 4%	32 ± 7%	22 ± 4%
Plasmids (ratio 1:4)				
<i>ras</i> + pMNC*	0 ± 1%	33 ± 4%	23 ± 3%	44 ± 3%
<i>ras</i> + p301-2 [†]	61 ± 5%	48 ± 5%	35 ± 7%	17 ± 4%
<i>ras</i> + p301-1 [‡]	84 ± 4%	67 ± 4%	25 ± 6%	7 ± 3%

Cells were transfected and G418-resistant (400 µg ml⁻¹) colonies were morphologically examined according to the criteria described in Fig. 2. Percentages are calculated for two experiments with ≥200 (a) or ≥400 (b) colonies per transfection.

* Vector control.

[†] Antisense orientation.

[‡] Sense orientation.

[§] The efficiency of inhibition of *ras* transformation is given as percentage reduction in the number of transformed colonies.

were generally smaller and grew slower than sense or vector control colonies. Ten out of ten antisense colonies showed barely detectable levels of Raf-1 protein, whereas levels in sense control clones were unchanged (data not shown). An alternative approach to Raf-1 inhibition used inactive mutants^{13,16}. A truncated Raf-1 protein (HCR) corresponding to conserved region 1 reduced colony numbers fourfold. A kinase-defective Raf-1 mutant protein, craf301 (plasmid p301-1), was even more efficient, decreasing colony yield about sevenfold¹⁶. The surviving colonies from these experiments could not be maintained as stable cell lines. We therefore turned to *raf*-transformed cell-lines where morphological reversion and inhibition of proliferation could be studied. p301 constructs were transfected into 208-F12 fibroblasts which overexpress a transforming mouse Raf-1 protein¹⁷. p301-2 caused partial or complete reversion of the transformed phenotype in approximately half the transfectants. Reversion correlated with loss of anchorage-independent growth. Again, p301-1 was more efficient than p301-2 (Fig. 2a). These clones were unstable, but cell-lines sufficiently stable for biochemical analysis were obtained after

‡ To whom correspondence should be addressed.

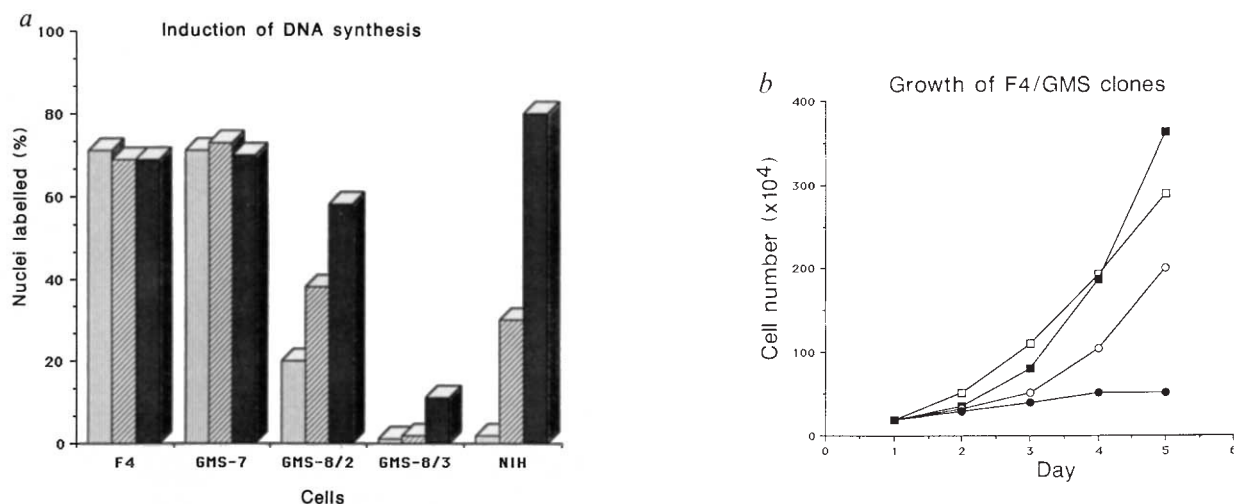


FIG. 3 Mitogen responsiveness and proliferative capacity of Raf-depleted cells. **a**, DNA synthesis induced by serum or TPA in serum-starved cells is depicted as the number of nuclei incorporating ^3H -thymidine. **b**, Long term growth curves GMS-7 is a pool of 10 clones transfected with sense DNA, GMS-8/2 and GMS-8/3 are flat clones reverted with antisense DNA. **a**, $\square$, starved cells; $\square$, TPA-induced cells; $\blacksquare$, SERUM-induced cells. **b**, $\square$, F4; $\circ$, GMS-7b; $\bullet$, GMS-8/3. **METHODS**. 10^4 cells were plated on cover slips and serum-starved for 24 h

before incubation with 20% foetal calf serum (Gibco) or 100 ng ml $^{-1}$ TPA (Sigma). 14 h after addition of mitogens, cells were labelled with 1 $\mu\text{Ci ml}^{-1}$ ^3H -thymidine for 5 h. Cells were counter-stained with Giemsa and labelled nuclei were counted. For long-term growth curves, 10^5 cells were seeded in six-well plates in DMEM medium supplemented with 10% FCS. Each day, one well was trypsinized and counted with a Coulter cell counter. All determinations were performed in triplicate.

pGMS transfection of *v-ras* transformed cells, F4 (ref. 18). Neither pMNC nor the control plasmid GMS-7 was effective, whereas the antisense construct, GMS-8, completely or partially reverted F4 (Fig. 2b). Reduction of *raf* mRNA (data not shown) and protein levels correlated with the extent of reversion (Fig. 2c). In one clone, GMS-8/3 (marked f* in Fig. 2c), *raf* protein expression was undetectable. These cells grew extremely slowly, arresting at 50–60% confluency, and eventually died.

To measure the effects of *raf*-protein depletion on the mitogen response, we determined the ability of serum and TPA to induce DNA synthesis in serum-starved cells (Fig. 3a). F4 and GMS-7 cells synthesize DNA independently of mitogens. Constitutive DNA synthesis was diminished in GMS-8/2, which retained an inducible response similar to NIH/3T3 cells. GMS-8/3 was completely blocked in constitutive and TPA-inducible DNA replication. Serum-stimulation of GMS-8/3 was reduced sevenfold, and long-term growth was also severely diminished (Fig. 3b).

There is ample evidence for control by *ras* of the signal flow from membrane receptor systems^{14,15,19–22}. We thus transfected *v-Ki-ras*-transformed NIH/3T3 cells with the p301 plasmids (Table 1a). p301-1 and p301-2 reduced neomycin (neo)-resistant colony yield to a similar degree as in NIH/3T3 cells (Fig. 1), suggesting that Raf-1 is required for proliferation of *ras*-trans-

formed cells. Morphological reversion of established *ras*-transformed cells was less efficient than of *raf*-transformed cells (Fig. 2). To test the effect of *raf*-inhibition on the initiation of *ras*-transformation, a constant amount of *v-Ha-ras* (pSV2neo/*ras*, (ref. 24) plasmid was co-transfected with an equal or four-molar amount of the p301 vectors (Table 1b). Although the neomycin resistance of pMNC-based plasmids accounted for a background of flat neo r colonies that presumably did not express pSV2neo/*ras*, transfection with p301 vectors markedly increased the number of morphological revertants at the expense of transformed colonies. The inhibition was dose-dependent and almost complete at four-molar excess of p301-1.

We conclude that in NIH/3T3 cells Raf-1 kinase functions downstream of membrane receptors and *ras* proteins and is essential for growth-induction by serum factors and protein kinase C. Membrane receptor systems can now be examined individually for Raf-1-dependence by inhibition with the blocking constructs described here. Furthermore, the proposed position of *raf* in the communication pathway between cell membrane and nucleus makes *raf* an attractive target for the design of novel antiproliferative agents, especially as our data show that *raf* inhibition is dominant over transformation by *ras* and, by implication, by other non-nuclear oncogenes. $\square$

Received 24 October; accepted 12 December 1990.

- Morrison, D. K. *et al.* *Cell* **58**, 649–657 (1989).
- App, H. *et al.* *Molec. cell. Biol.* (in the press).
- Blackshear, P. J., Haupt, D. M., App, H. & Rapp, U. R. *J. biol. Chem.* **265**, 12131–12134 (1990).
- Kovachina, K. S. *et al.* *J. biol. Chem.* **265**, 12115–12118 (1990).
- Turner, B. C. *et al.* *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Carroll, M. P., Clarke-Lewis, I., Rapp, U. R. & May, W. S. *J. biol. Chem.* **265**, 19812–19817 (1990).
- Baccarini, M., Sabatini, D. M., App, H., Rapp, U. R. & Stanley, E. R. *EMBO J.* **9**, 3649–3657 (1990).
- Morrison, D. K., Kaplan, D. R., Rapp, U. R. & Roberts, T. M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8855–8859 (1988).
- Rapp, U. R. *et al.* in *Cold Spring Harbor Symposium on Quantitative Biology*, Vol. LIII, 173–184 (Cold Spring Harbor Laboratory, New York, 1988).
- Siegel, J. N., Klausner, R. D., Rapp, U. R. & Samelson, L. E. *J. biol. Chem.* **265**, 18472–18480 (1990).
- Storm, S. M., Cleveland, J. L. & Rapp, U. R. *Oncogene* **5**, 345–351 (1990).
- Smith, M. R., Heidecker, G., Rapp, U. R. & Kung, H. *Molec. cell. Biol.* **10**, 3828–3833 (1990).
- Rapp, U. R., Cleveland, J. L., Bonner, T. I. & Storm, S. M. in *The Oncogene Handbook* (eds E. P. Reddy, A. M. Skalka, and T. Curran) 213–253 (Elsevier Science, The Netherlands, 1988).
- Smith, M. R., DeGudicibus, S. J. & Stacey, D. W. *Nature* **320**, 540–543 (1986).

- Noda, M., Selinger, Z., Scolnik, E. M. & Bassin, R. H. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5602–5606 (1983).
- Heidecker, G. *et al.* *Molec. cell. Biol.* **10**, 2503–2512 (1990).
- Schultz, A. M., Copeland, Oroszlan, S. & Rapp, U. R. *Oncogene* **2**, 187–193 (1988).
- Rapp, U. R. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **80**, 4218–4222 (1983).
- Mulcahy, L. S., Smith, M. R. & Stacey, D. W. *Nature* **313**, 241–243 (1985).
- Yu, C. L., Tsai, M. H. & Stacey, D. W. *Cell* **52**, 63–71 (1988).
- Cai, H., Szeberenyi, J. & Cooper, G. M. *Molec. cell. Biol.* **10**, 5314–5323 (1990).
- Szeberenyi, J., Cai, H. & Cooper, G. M. *Molec. cell. Biol.* **10**, 5324–5332 (1990).
- Huleihel, M. *et al.* *Molec. cell. Biol.* **6**, 2655–2662 (1986).
- Clanton, D. J., Lu, Y., Blair, D. G. & Shih, T. Y. *Molec. cell. Biol.* **7**, 3092–3097 (1987).
- Bonner, *et al.* *Nucleic Acids Res.* **14**, 1009–1015 (1986).
- Kolch, W. *et al.* *Oncogene* **5**, 713–720 (1990).

ACKNOWLEDGEMENTS. We thank Brian Seed for pMNC, Thomas Shih for pSV2neo/*ras*, Howard Young for *v-Ki-ras* transformed NIH/3T3 cells, and Mark Smith for help with the ^3H -thymidine incorporation assays.

Localization of p53, retinoblastoma and host replication proteins at sites of viral replication in herpes-infected cells

D. Wilcock & D. P. Lane*

Imperial Cancer Research Fund, Clare Hall Laboratories,
South Mimms, Potters Bar, Hertfordshire EN6 3LD, UK

* Present address: Cancer Research Campaign Laboratories, University of
Dundee, Dundee DD1 4HN, UK

REPLICATION of DNA occurs at discrete sites in eukaryotic cell nuclei, where replication proteins are clustered into large complexes, or 'replicases'¹⁻³. Similarly, viral DNA replication is a highly structured process, notably in herpes simplex virus type-1 (HSV-1; reviewed in ref. 4) in which large globular 'replication compartments' containing the viral replication machinery exist. Replicating cellular DNA redistributes to these compartments upon HSV-1 infection⁵. We have now used antibodies raised against several cellular proteins to detect changes in their subnuclear localization on HSV-1 infection. We found that various proteins involved in cellular DNA replication move to sites of viral DNA synthesis, whereas a selection of non-replication proteins do not. The retinoblastoma protein and p53 (the products of two putative anti-oncogenes^{6,7}) relocate to the same sites as known DNA replication proteins, suggesting that they may be associated with DNA replication complexes in normal, uninfected cells.

Infected cell protein 8 (ICP8), the HSV-1 major DNA binding protein, is associated with viral DNA in replication compartments⁸; thus antiserum to ICP8 was used as a convenient marker for viral replication sites. Immunofluorescent double-labelling, using both confocal and conventional microscopy, showed that the cellular single-stranded DNA binding protein SSB (also known as RF-A, or RP-A⁹) proliferating cell nuclear antigen (PCNA, an accessory factor to polymerase δ)^{10,11}, the retinoblastoma protein (Rb) and p53 all colocalize with ICP8 in HSV-1-infected cells (Fig. 2*a-d*), in contrast to their granular distribution in the nuclei of uninfected cells (Fig. 1). The colocalization of SSB is complete, whereas p53 is found exclusively at certain sites and some Rb fails to enter the compartments.

The results are clearer when cells are infected in the presence of phosphonoacetate (PAA), a specific inhibitor of the HSV-1 DNA polymerase¹². ICP8 is now present in smaller, more numerous regions, called 'prereplicative sites'⁸. On PAA treatment, SSB is always found at these sites (Fig. 3*a*), whereas PCNA, Rb and p53 generally show only partial colocalization with ICP8 (Fig. 3*b-d*). This indicates that SSB may redistribute

TABLE 1 Status of host proteins in cells infected with HSV-1 (no PAA treatment)

Colocalization with ICP8	No colocalization with ICP8
SSB	snRNP
PCNA	c-Myc
Rb	lamin B
p53	p68
DNA ligase 1	nucleophosmin (B23)
pol α	Ki67

SSB, PCNA, Rb, p53 and p68 are discussed in the text. SSB was detected with antibodies 70C and 34A⁹. PCNA was detected with the PC2, PC4, and PC1 antibodies³⁴. Rb was detected with antibodies Rb-PMG3-245 (Pharmin-gen, San Diego)²⁸ and IF/8 (from J. Bartek). p53 was detected with antibodies PAb421³⁵ and PAb1801³⁶. p68 was detected with the PAb204 antibody¹². DNA ligase 1 is the replicative ligase and was stained with a polyclonal antibody³⁷. Antibody SJK-132³⁸ was used to detect DNA polymerase α . The small nuclear ribonucleoprotein particle (snRNP) was detected with antibody Y12 (from J. Steitz, Yale)³⁸, and c-Myc was visualized with antibody 9E10³⁹. Antibody LN43 (from B. Lane) was used to detect lamin B, a structural nuclear protein. The nucleolar protein B23 was detected with a monoclonal antibody (from P. K. Chan, Texas Medical Center)⁴⁰. Ki67 is also a nucleolar protein, present in proliferating cells⁴¹ (antibody from Immunotech, France).

by a mechanism different from that used for PCNA, Rb and p53 relocation, or perhaps binding sites for the latter proteins become saturated. The extent of colocalization of these latter proteins varies from cell to cell (Fig. 3*c-d* shows particularly clear-cut examples of colocalization). Thus the accumulation of host proteins at these sites may occur in distinct stages.

Infection of serum-starved cells (5 days in 0.1% bovine serum) produced similar results, suggesting that redistribution of host proteins may occur during natural infections of resting cells. Control staining experiments showed no cross-reaction between the fluorescent conjugates used. Immunoblotting of total infected cell extracts showed that none of the antibodies used cross-reacted with HSV-1 proteins. In addition, the anti-ICP8 antibody did not cross-react with any of the cellular proteins studied. Several monoclonal antibodies recognizing discrete epitopes were used to detect SSB, PCNA, Rb and p53 proteins (Table 1).

The redistribution is not simply a nonspecific aggregation of nuclear proteins by the virus, because other host nuclear proteins, such as p68 (ref. 13) (Figs 2*e*, 3*e*), lamin B, nucleophosmin and Ki67 (Table 1) do not redistribute on infection. Small ribonuclearprotein particles¹⁴ and c-Myc (Table 1) show a different pattern of redistribution on infection, forming clusters peripheral to the compartments. Only a specific subset of viral proteins are present in replication compartments^{15,16}. Hoescht staining showed no gross movement of cellular DNA into replication compartments, but rather an exclusion of cellular chromatin from these areas, as seen previously⁵.

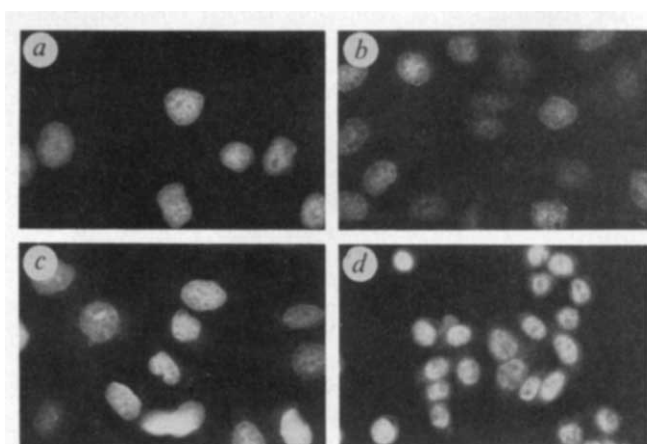


FIG. 1 Cell staining of uninfected cells. Immunofluorescence pictures of monkey CV-1 cells (*a, b, c*), or a p53-overproducing breast cancer cell line, MDA-MB-468 (*d*)³¹. Cells stained with *a*, 70C, anti-SSB; *b*, PC4, anti-PCNA; *c*, IF/8, anti-Rb; *d*, PAb421, anti-p53.

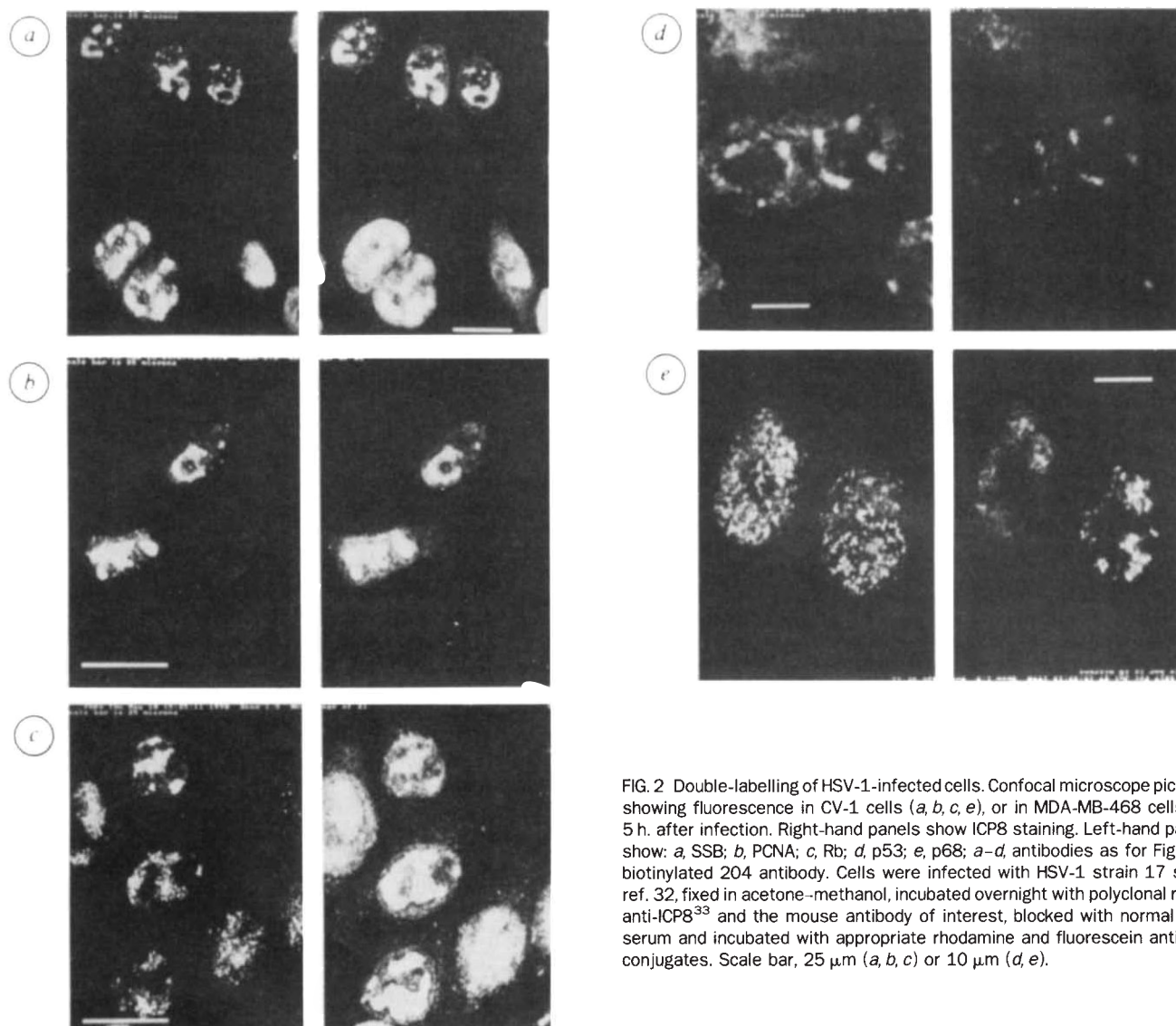


FIG. 2 Double-labelling of HSV-1-infected cells. Confocal microscope pictures showing fluorescence in CV-1 cells (*a, b, c, e*), or in MDA-MB-468 cells (*d*), 5 h. after infection. Right-hand panels show ICP8 staining. Left-hand panels show: *a*, SSB; *b*, PCNA; *c*, Rb; *d*, p53; *e*, p68; *a–d*, antibodies as for Fig. 1 *e*, biotinylated 204 antibody. Cells were infected with HSV-1 strain 17 syn + ref. 32, fixed in acetone-methanol, incubated overnight with polyclonal rabbit anti-ICP8³³ and the mouse antibody of interest, blocked with normal goat serum and incubated with appropriate rhodamine and fluorescein antibody conjugates. Scale bar, 25 μm (*a, b, c*) or 10 μm (*d, e*).

Table 1 groups the proteins studied in terms of their distribution in infected cells. It can be seen that, at least on this basis, Rb and p53 behave as proteins associated with replication complexes. Although Rb, p53 and ICP8 colocalize, no physical association between the proteins was detected by western blotting following immunoprecipitation, suggesting that any interaction must be either weak or indirect.

The redistribution of replicating cellular DNA on HSV-1 infection suggested that replication proteins might also be moved by the virus. Functional ICP8 is required for the redistribution of replicating cellular DNA⁵, implying that this protein may be involved in moving replication complexes.

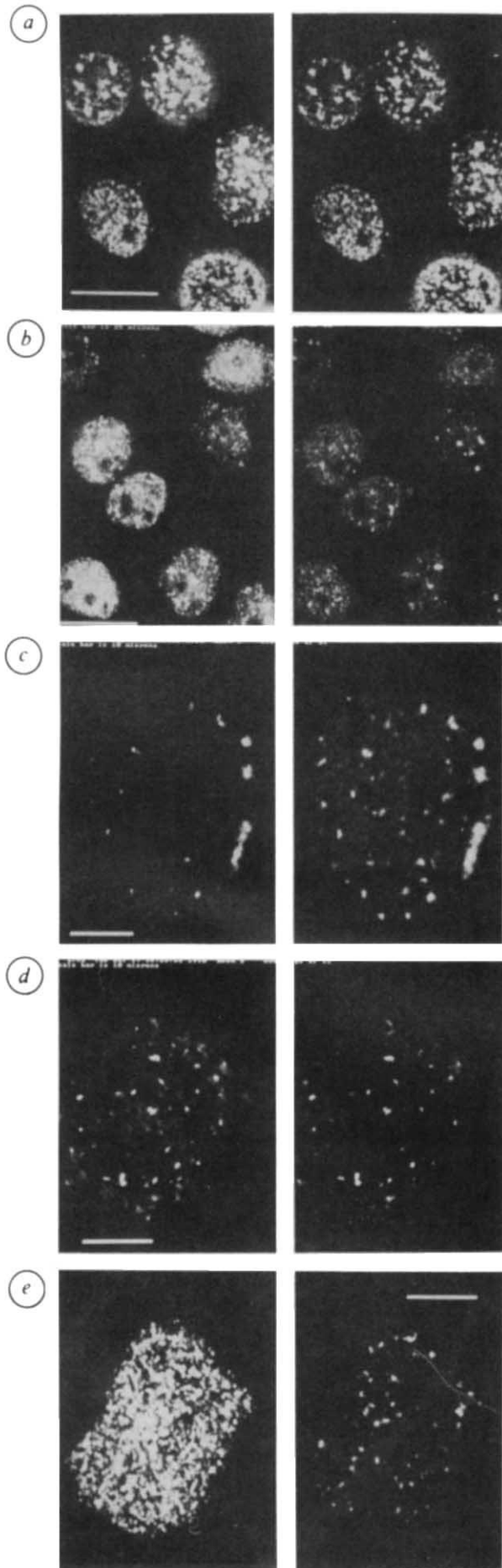
Whether some, or all, of the redistributed proteins play a role in HSV-1 replication is unclear. Seven viral genes essential for DNA replication have been identified by transient transfection studies¹⁷ and analysis of HSV mutants. These genes encode a DNA polymerase and accessory factor, ICP8, a three-subunit helicase-primase complex¹⁸, and an origin-binding protein¹⁹. Thus there seems to be an element of functional redundancy between the cellular and viral proteins in replication compartments. Perhaps some component of the replicase is targeted for transport by the virus (perhaps through ICP8) and as a result the whole complex redistributes to a replication compartment, where certain replication proteins may be used

by the virus, while others remain only passively.

Recompartmentalization of nuclear proteins is also a feature of adenovirus replication, where certain host nucleolar proteins are moved to viral 'replication factories'²⁰. Both herpesvirus and adenovirus can act as helper viruses for minute virus of mice²¹, which can normally replicate only in S-phase cells²². The helper-virus function may be to make the host cell competent for replication by recompartmentalizing nuclear proteins²³.

Other proteins from DNA viruses interact with p53 and Rb; simian virus 40 (SV40) large T antigen, adenovirus E1b and papillomavirus E6 proteins all bind to p53^{24–27}, whereas large T, adenovirus E1a and papillomavirus E7 bind to Rb^{28–30}. It may be that the function of p53 and Rb needs to be altered for these viruses to overcome quiescence in the host cell, thus creating an environment in which viral replication can occur. This model relates to the proposed roles of p53 and Rb as tumour suppressors: elimination of p53 and Rb function may

FIG. 3 Double-labelling of cells infected in the presence of 400 $\mu\text{g ml}^{-1}$ PAA, 5 h after infection. Left panels: *a*, SSB; *b*, PCNA (PC2); *c*, Rb; *d*, p53; *e*, p68. Right panels show ICP8. Antibodies, cells and methods as described in Fig. 2. Scale bar, 25 μm (*a, b*) or 10 μm (*c, d, e*). ▶



lead to a cell which permits permanent S phase, thus facilitating viral replication and tumour development. Dissection of the HSV-1 infection system may shed light on the functions and cellular targets of these two anti-oncogenes. □

Received 31 August; accepted 6 December 1990.

1. Nakamura, H., Morita, T. & Sato, C. *Exp Cell Res* **165**, 291-297 (1986).
2. Nakayasu, H. & Berezney, R. *J Cell Biol* **108**, 1-11 (1989).
3. Mills, A. D. et al. *J. cell. Sci.* **94**, 471-477 (1989).
4. Knipe, D. M. *Adv. Virus Res.* **37**, 85-123 (1990).
5. de Bruyn Kops, A. & Knipe, D. M. *Cell* **55**, 857-868 (1988).
6. Lee, W.-H. et al. *Nature* **329**, 642-645 (1987).
7. Lane, D. P. & Benichou, S. *Genes Dev.* **4**, 1-8 (1990).
8. Quinlan, M. P., Chen, L. B. & Knipe, D. M. *Cell* **36**, 857-868 (1984).
9. Kenny, M. K., Schlegel, U., Furneaux, H. & Hurwitz, J. *J. biol. Chem.* **265**, 7693-7700 (1990).
10. Bravo, R., Frank, R., Blundell, P. A. & MacDonald-Bravo, H. *Nature* **326**, 515-517 (1987).
11. Prelich, G. et al. *Nature* **326**, 517-520 (1987).
12. Mao, J. C. H. & Robishaw, E. E. *Biochemistry* **14**, 5475-5479 (1975).
13. Iggo, R. D. & Lane, D. P. *EMBO J.* **8**, 1827-1831 (1989).
14. Martin, T. E., Barghusen, S. C., Leser, G. P. & Spear, P. G. *J. Cell Biol.* **105**, 2069-2082 (1987).
15. Knipe, D. M., Senechek, D., Rice, S. A. & Smith, J. L. *J. Virol.* **61**, 276-284 (1987).
16. Olivo, P. D., Nelson, N. J. & Challberg, M. D. *J. Virol.* **63**, 196-204 (1989).
17. Wu, C. A., Nelson, N. J., McGeoch, D. J. & Challberg, M. D. *J. Virol.* **62**, 435-443 (1988).
18. Crute, J. J. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2187-2191 (1989).
19. Weir, H. M., Calder, J. M. & Stow, N. D. *Nucleic Acids Res.* **17**, 1409-1425 (1989).
20. Walton, T. H., Moen Jr, P. T., Fox, E. & Bodnar, J. W. *J. Virol.* **63**, 3651-3660 (1989).
21. Berns, K. I. & Hauswirth, W. W. in *Replication of Mammalian Parvoviruses* (eds Ward, D. C. & Tattersall, P.) 13-32 (Cold Spring Harbor, New York, 1978).
22. Wolter, S., Richards, R. & Armentrout, R. W. *Biochim. biophys. Acta* **607**, 420-431 (1980).
23. Fox, E., Moen, P. T. Jr. & Bodnar, J. W. *Virology* **176**, 403-412 (1990).
24. Lane, D. P. & Crawford, L. V. *Nature* **278**, 261-263 (1979).
25. Linzer, D. I. H. & Levine, A. J. *Cell* **17**, 43-52 (1979).
26. Sarnow, P., Ho, Y. S., Williams, J. & Levine, A. J. *Cell* **28**, 387-394 (1982).
27. Werness, B. A., Levine, A. J. & Howley, P. M. *Science* **248**, 76-79 (1990).
28. DeCaprio, J. A. et al. *Cell* **54**, 275-283 (1988).
29. Whyte, P. et al. *Nature* **334**, 124-129 (1988).
30. Dyson, N., Howley, P. M., Munger, K. & Harlow, E. *Science* **243**, 934-936 (1989).
31. Bartek, J., Iggo, R., Gannon, J. & Lane, D. P. *Oncogene* **5**, 893-899 (1990).
32. Brown, S. M., Ritchie, D. A. & Subak-Sharpe, J. H. *J. gen. Virol.* **18**, 329-346 (1973).
33. Powell, K. L., Littler, E. & Purifoy, D. J. M. *J. Virol.* **39**, 894-902 (1981).
34. Waseem, N. H. & Lane, D. P. *J. Cell Sci.* **96**, 121-129 (1990).
35. Harlow, E., Crawford, L. V., Pim, D. C. & Williamson, N. M. *J. Virol.* **39**, 861-869 (1981).
36. Banks, L., Matlashewski, G. & Crawford, L. V. *Eur. J. Biochem.* **159**, 529-534 (1986).
37. Tomkinson, A. E., Lasko, D. D., Daly, G. & Lindahl, T. *J. biol. Chem.* **265**, 12611-12617 (1990).
38. Tanaka, S., Hu, S.-Z., Wang, T. S. F. & Korn, D. *J. biol. Chem.* **257**, 8386-8390 (1982).
39. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. *Nature* **283**, 220-224 (1980).
40. Evan, G. I., Lewis, G. K., Ramsay, G. & Bishop, J. M. *Molec. cell. Biol.* **5**, 3610-3616 (1985).
41. Ochs, R., Lischwe, M., O'Leary, P. & Busch, H. *Exp Cell Res.* **146**, 139-149 (1983).
42. Gerdes, J., Schwab, U., Lemke, H. & Stein, H. *Int. J. Cancer* **31**, 13-20 (1983).

ACKNOWLEDGEMENTS. We thank K. Powell (Wellcome, UK) for the anti-ICP8 serum, N. Stow (MRC, Glasgow) for the HSV strain, and the CRC Molecular Embryology group in Cambridge for use of their confocal microscope.

Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG

Shinya Shibutani, Masaru Takeshita
& Arthur P. Grollman

Department of Pharmacological Sciences, State University of New York at Stony Brook, Stony Brook, New York 11794-8651, USA

OXIDATIVE damage to DNA, reflected in the formation of 8-oxo-7-hydrodeoxyguanosine (8-oxodG)^{1,2}, may be important in mutagenesis, carcinogenesis and the ageing process^{3,4}. Kuchino *et al.* studied DNA synthesis on oligodeoxynucleotide templates containing 8-oxodG, concluding that the modified base lacked base pairing specificity and directed misreading of pyrimidine residues neighbouring the lesion⁵. Here we report different results, using an approach in which the several products of a DNA polymerase reaction can be measured. In contrast to the earlier report⁵, we find that dCMP and dAMP are incorporated selectively opposite 8-oxodG with transient inhibition of chain extension occurring 3' to the modified base. The potentially mutagenic insertion of dAMP is targeted exclusively to the site of the lesion. The ratio of dCMP to dAMP incorporated varies, depending on the DNA polymerase involved. Chain extension from the dA·8-oxodG pair was efficiently catalysed by all polymerases tested.

5' CCTTCXCTACTTTCCTCT
 TGAAAGGAGA 5'
 ATGAAAGGAGA 5'
 GATGAAAGGAGA 5'
 N GATGAAAGGAGA 5'

X = dG or 8-oxodG

FIG. 1 Structure of 8-oxodG and sequence of templates and primers. Oligodeoxynucleotides, including those containing 8-oxodG, were prepared by solid-state synthesis on a DuPont Coder 300 automated DNA synthesizer as reported elsewhere¹⁷, removed from the resin-support by treatment with concentrated ammonia containing 0.1 M 2-mercaptoethanol at 55 °C for 16 h, then purified by electrophoresis on 20% polyacrylamide gels (72 × 15 × 0.2 cm) in the presence of 7 M urea. Primers (100 pmol) were labelled with ³²P at the 5' terminus with 5 µl of bacteriophage T4 polynucleotide kinase (10 units µl⁻¹) in the presence of 5 µl [γ -³²P]ATP (10 Ci µl⁻¹)¹⁸, heated at 100 °C for 3 min and concentrated on Centricon filters after adding 500 µl distilled water. Labelled primer (50 pmol) was added to 75 pmol template DNA, heated at 56 °C for 15 min then annealed by standing at room temperature for 1 h and overnight at 4 °C.

Figure 1 lists nucleotide sequences of primers and template used in these experiments. In Fig. 2, 18-mers containing dC, dA, dG or dT at position 13 are distinguished by their relative electrophoretic mobility on 20% polyacrylamide gels (lanes 6–9). DNA synthesis on unmodified templates, catalysed by DNA polymerase α (pol α) (lane 1) and other polymerases (data not shown) led to the expected exclusive incorporation of dCMP at position 13. Using the DNA polymerases indicated in Fig. 2 and templates modified with 8-oxodG, 18-mers containing dA and dC were formed (lanes 2–5). The ratio of dC to dA for 18-mers synthesized by the Klenow fragment was 7:1; by intact pol I, 7:1 (data not shown), by pol β 4:1; by pol α 1:200; and by pol δ , 1:5 18-mers containing dG or dT were not detected. In experiments using the Klenow fragment or DNA pol α , the identity of dCMP and/or dAMP incorporated opposite the lesion was confirmed by Maxam–Gilbert sequence analysis⁶.

Kinetic parameters of base insertion and extension, catalysed by Klenow fragment and polymerase α , were measured during translesional synthesis (Table 1). These reactions measure addition of a single dNMP to various 3' termini under steady-state conditions. The Michaelis constant, K_m , and the maximum rate

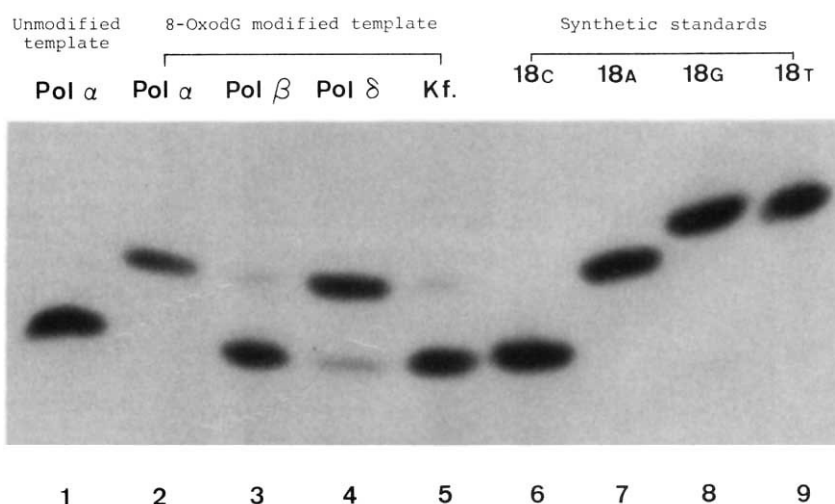
of the reaction, V_{max} , were determined experimentally; the frequency of dNMP insertion and extension opposite 8-oxodG was calculated relative to experiments using the unmodified template, as described by Mendelman *et al.*^{7,8}. These kinetic constants are highly reproducible. Using the Klenow fragment, the frequency of insertion (F_{ins}) for dCMP opposite 8-oxodG (0.35) is seven times higher than for dAMP. The frequency of extension (F_{ext}) from the dC·8-oxodG base pair is, however, 12 times lower than from dA·8-oxodG. In reactions catalysed by calf thymus pol α , F_{ins} of dAMP opposite the lesion (0.13) is seven times higher than for dCMP; the dA·8-oxodG pair is efficiently extended. Similar results were obtained using human pol α (not shown). We conclude from these experiments that dCTP and dATP compete for insertion on DNA templates containing 8-oxodG. The polymerases tested vary greatly in their respective preference for inserting dA and dC opposite the lesion and extend preferentially the dA·8-oxodG base pair.

Kuchino *et al.*⁵ reported misreading of pyrimidine residues located adjacent to 8-oxodG. Using the Klenow fragment and similar experimental conditions, we measured the relative insertion of dNMPs immediately 3' to the lesion on a series of templates where the 5' flanking base was dC, dG, dT or dA. In all cases, dG was inserted exclusively opposite the 3' dC residue (not shown). Misincorporation of nucleotides opposite pyrimidines adjacent to the lesion was not observed even when sequences of the template used were identical to those reported by Kuchino *et al.*⁵.

We also tested the proofreading ability of Klenow fragment (not shown) and intact DNA polymerase I to replace bases inserted opposite 8-oxodG at the 3' terminus (Fig. 3). Although dA was readily excised and replaced by dC in dA·dG mismatches, neither dA nor dC was replaced when paired with 8-oxodG. Both base pairs were readily extended, however. Apparently, the 3'→5' exonuclease proof-reading function of DNA polymerase I does not recognize dA·8-oxodG or dC·8-oxodG as a substrate.

Fidelity of DNA polymerases is primarily achieved through selective insertion of dNTPs opposite the lesion, proof-reading by 3'→5' exonuclease and extension of the newly formed base pair^{7–10}. This process is illustrated by our experiments with the Klenow fragment in which dAMP is inserted opposite 8-oxodG at a lower frequency than dCMP, forming a base pair that is more efficiently extended than dC·8-oxodG. In this case, proof-reading does not affect the composition of the final product of translesional synthesis, which is determined primarily by F_{ins} .

FIG. 2 Incorporation of nucleotides opposite 8-oxodG. Chain extension reactions were carried out in a solution (10 µl) containing DNA polymerase, buffer, salts, four dNTPs (100 µM each) and an unmodified 18-mer (lane 1) or 8-oxodG-modified 18-mer template primed with ³²P-labelled 10mer (0.05 µM), except in the case of DNA pol δ , in which a ³²P-labelled 12-mer was used. Sources of DNA polymerase are as follows: HeLa DNA pol α and pol β ^{19,20} (Dr S. Wilson), calf thymus pol δ ^{21,22} (Dr K. Downey), *Escherichia coli* pol I (Pharmacia) and the cloned Klenow fragment (Kf) (International Biotechnologies, Inc.) HeLa DNA pol α (1 unit) was incubated at 25 °C for 1 h in 50 mM Tris-HCl buffer (pH 8.0) containing 10 mM MgCl₂, 20 mM ammonium sulphate, 2 mM dithiothreitol (DTT) and BSA (0.5 µg µl⁻¹). HeLa DNA pol β (1 unit) was incubated for 1 h at 25 °C in the same buffer used for pol α , omitting ammonium sulphate. Calf thymus DNA pol δ (0.7 units) was incubated at 30 °C for 2 h in 50 mM Tris-HCl buffer (pH 6.5) containing 10 mM MgCl₂, 2 mM DTT, BSA (0.04 µg µl⁻¹) and 100 ng human PCNA (proliferating cell nuclear antigen) cloned in *E. coli*, provided by Dr P. Fisher. The Klenow fragment (0.05 units) was incubated at 20 °C for 1 h in 50 mM Tris-HCl buffer (pH 8.0) containing 8 mM MgCl₂ and 5 mM 2-mercaptoethanol. Reaction mixtures were subjected to 20%



polyacrylamide gel electrophoresis. Mobilities of the 18-mer reaction products (lanes 1–5), are compared with those of synthetic oligodeoxynucleotides containing dC (lane 6), dA (lane 7), dG (lane 8) or dT (lane 9) at position 13.

TABLE 1 Kinetic parameters of insertion and extension reactions catalysed by Klenow fragment and DNA polymerase α

	dNTP ↓G— 12-mer 5'-CCTTCXC— $V_{\max}$ (% min ⁻¹)				dGTP ↓NG— 13-mer 5'-CCTTCXC— $V_{\max}$ (% min ⁻¹)		
	K_m (μ M)		F_{ins}		K_m (μ M)		F_{ext}
(a)							
	2.04	14.5	1.00		0.63	24.8	1.00
	ND	—	—		ND	—	—
	6.62	16.3	0.35		8.41	10.8	0.033
	28.3	9.5	0.05		1.00	19.4	0.50
(b)							
	5.54	4.2	1.00		11.7	4.9	1.00
	1,410	0.17	1.56×10^{-4}		4,990	0.05	2.47×10^{-5}
	70.5	0.96	0.018		1,160	0.46	9.45×10^{-4}
	42.3	4.1	0.13		26.1	2.0	0.19

a. Kinetics of insertion (0.1 μ M–10 mM dNTP) using the Klenow fragment (0.005 units) were determined at 20 °C; reaction mixtures were incubated for 90 s in 10 μ l Tris-HCl buffer (pH 8.0) containing template DNA primed with ³²P-labelled 12-mer (0.05 μ M), as described in the legend to Fig. 2. Experimental conditions are analogous to those reported by Mendelman *et al.*^{7,8}; all reactions were linear with time and data reported are an average of three separate experiments. Kinetics of chain extension (0.05 μ M–10 mM dGTP) were determined in reaction mixtures incubated at 20 °C for 90 s in 10 μ l Tris-HCl buffer (pH 8.0) containing template DNA primed with ³²P labelled 13-mer containing dC or dA (0.05 μ M). b. Kinetics of insertion (1.0 μ M–10 mM dNTP) using calf thymus pol α (1 unit, Molecular Biology Resources Inc.) were measured at 25 °C in reactions incubated for 5 min (dC·dG pairs), 10 min (dA·8-oxodG), 15 min (dC·8-oxodG), and 30 min (dA·dG). Kinetics of extension (0.4 μ M–10 mM dGTP) were determined in reactions conducted at 25 °C for 5 min (dC·dG), 10 min (dA·8-oxodG), and 30 min (dC·8-oxodG, dA·dG). Bands were cut from the gels and radioactivity was determined by liquid scintillation counting. Values for K_m and $V_{\max}$ were obtained from Hanes–Woolf plots. Insertion and extension frequencies (F) were determined relative to dC·dG according to equations described by Mendelman *et al.*^{7,8} where $F = (V_{\max}/K_m) [\text{wrong pair}] / (V_{\max}/K_m) [\text{right pair}]$ with 'wrong pair' defined either as a mismatch or any pair involving 8-oxodG. G:X, dG; G:X, 8-oxodG.

It seems that DNA polymerases differ significantly in their ability to incorporate specific dNMPs on the same DNA template. 8-oxodG occurs predominantly as the 6,8-diketo tautomer^{11–13} and assumes the *syn* conformation as a nucleoside^{13,14} or when located opposite dA in duplex DNA¹⁵.

8-oxodG assumes the *anti* conformation when pairing with dC (E. Ohtsuka, personal communication). During DNA replication or repair, enzymes could encounter the lesion in either conformation.

Our results differ significantly from those described by Kuchino *et al.*⁵ who reported that templates modified with 8-oxodG directed insertion of dA, dT, dC and dG opposite the lesion with almost equal frequency and caused misreading at neighbouring positions, an effect that depended on the sequence context. They reported also that the rate of DNA synthesis on 8-oxodG templates was not retarded. The discrepancy may be explained by the DNA polymerase used for dideoxynucleoside sequencing analysis by Kuchino *et al.*⁵. We found that Klenow fragment was not suitable for this particular procedure and used instead T7 DNA polymerase (Sequenase Version 2.0, USB), which lacks an exonuclease function. Under these conditions, only dA and dC were incorporated opposite the lesion; a result which was confirmed by direct Maxam–Gilbert sequencing analysis and by the relative electrophoretic migration of the reaction products. Targeted $G \rightarrow T$ transversions predicted by these *in vitro* experiments have emerged as the predominant base substitution induced *in vivo* by plasmids and phagemids modified with 8-oxodG, as studied by site-specific mutagenesis techniques^{16,23}.

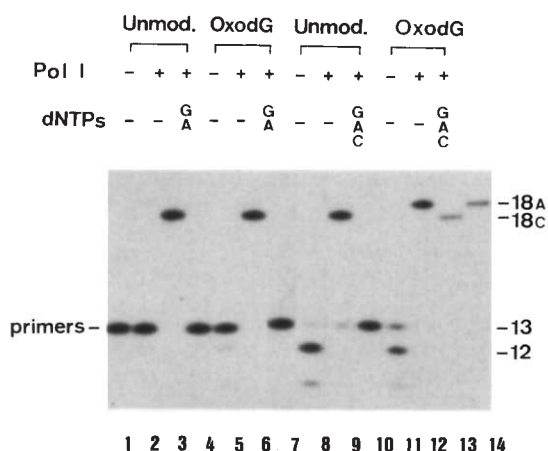


FIG. 3 Proof-reading of the 3' terminus by DNA pol I. Unmodified and 8-oxodG-modified 18-mer templates primed with a 13-mer containing dC (lanes 1–6) or dA (lanes 7–12) opposite the modified base were incubated at 20 °C for 1 h in the presence (lanes 2, 5, 8, 11) or absence (lanes 1, 4, 7, 10) of DNA pol I (0.05 unit) or with pol I and 100 μ M of each dNTP (lanes 3, 6, 9, 12) using described in the legend to Fig. 2. Lanes 13 and 14 contain synthetic 18-mers containing dC and dA, respectively, at position 13.

Received 22 August; accepted 15 November 1990.

1. Kasai, H. & Nishimura, S. *Nucleic Acid Res.* **12**, 2137–2145 (1984).
2. Shigenaga, M. K., Gimeno, C. J. & Ames, B. N. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9697–9701 (1989).
3. Harman, D. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7124–7128 (1981).
4. Ames, B. N. *Science* **221**, 1256–1264 (1983).

5. Kuchino, Y. et al. *Nature* **327**, 77-79 (1987).
6. Maxam, A. M. & Gilbert, W. *Meth. Enzymol.* **65**, 499-560 (1980).
7. Mendelman, L. V., Petruska, J. & Goodman, M. F. *J. biol. chem.* **264**, 14415-14423 (1989).
8. Mendelman, L. V., Petruska, J. & Goodman, M. F. *J. biol. Chem.* **265**, 2338-2346 (1990).
9. Loeb, L. A. & Kunkel, T. A. *A. Rev. Biochem.* **52**, 429-457 (1982).
10. Kunkel, T. A. *Cell* **53**, 837-840 (1988).
11. Kasai, H., Nishimura, S., Toriumi, Y., Itai, A. & Iitaka, Y. *Bull. Chem. Soc. Jpn* **60**, 3799-3800 (1987).
12. Aida, M. & Nishimura, S. *Mut. Res.* **192**, 83-89 (1987).
13. Culp, S. J., Cho, B. P., Kadlubar, F. F. & Evans, F. E. *Chem. Res. Toxicol.* **2**, 416-422 (1989).
14. Uesugi, J. & Ikehama, M. *J. Am. Chem. Soc.* **99**, 3250-3253 (1977).
15. Kouchakdjian, M. et al. *J. Biochemistry* (in the press).
16. Wood, M. L., Dizdaroğlu, M., Gajewski, E. & Essigmann, J. M. *Biochemistry* **29**, 7024-7032 (1990).
17. Bodepudi, V., Iden, C. R. & Johnson, F. *Nucleosides and Nucleotides* (in the press).
18. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning* 13.73-13.77 (Cold Spring Harbor Press, New York, 1989).
19. Abbots, J. et al. *Biochemistry* **27**, 901-909 (1988).
20. Karawya, E. M. & Wilson, S. H. *J. biol. Chem.* **257**, 13129-13134 (1982).
21. Lee, M. Y. W. T., Tan, C.-K., Downey, K. M. & So, A. G. *Biochemistry* **23**, 1906-1913 (1984).
22. Tan, C.-K., Castillo, C., So, A. G. & Downey, K. M. *J. biol. Chem.* **261**, 12310-12316 (1986).
23. Moriya, M. et al. *Mut. Res.* (in the press).

ACKNOWLEDGEMENTS. We thank R. Rieger for preparing the modified oligodeoxynucleotides, S. H. Wilson (NIH) for gifts of human polymerases α and β and P. Fisher for human PCNA. These studies were supported by the National Cancer Institute and the National Institute of Environmental Health Sciences.

RNA editing by cytidine insertion in mitochondria of *Physarum polycephalum*

R. Mahendran, M. R. Spottswood & D. L. Miller*

Cell and Molecular Biology Program, The University of Texas at Dallas, Box 830688, Richardson, Texas 75083-0688, USA

A COROLLARY of the central dogma of molecular biology is that genetic information passes from DNA to RNA by the continuous synthesis of RNA on a DNA template. The demonstration of RNA editing¹ (the specific insertion, deletion or substitution of residues in RNA to create an RNA with a sequence different from its own template) raised the possibility that in some cases not all of the genetic information for a trait resides in the DNA template. Two different types of RNA editing have been identified in mitochondria: insertional editing represented by the extensive insertion (and occasional deletion) of uridine residues in mitochondrial RNAs of the kinetoplastid protozoa²⁻⁴ and the substitutional editing represented by the cytidine to uridine substitutions in some plant mitochondria⁵⁻⁷. These editing types have not been shown to be present in the same organism and may have very different mechanisms. RNA editing of both types has been observed in non-mitochondrial systems⁸⁻¹⁷ but is not as extensive and may involve

still different mechanisms. Here we report the discovery of extensive insertional RNA editing in mitochondria from an organism other than a kinetoplastid protozoan. The mitochondrial RNA apparently encoding the α subunit of ATP synthetase in the acellular slime mould, *Physarum polycephalum*, is edited at 54 sites by cytidine insertion.

The mitochondrial DNA (mtDNA) of *P. polycephalum* is a circle of 60 kilobases (kb) and is extensively transcribed¹⁸. A 970-nucleotide complementary DNA with homology to *P. polycephalum* mtDNA was identified by screening a *Physarum* cDNA library with uniformly radiolabelled mtDNA. The location of the mtDNA homology was determined by radiolabelling the cDNA and hybridizing it with Southern blots of mtDNA digested with various restriction enzymes. The hybridization pattern indicated that the homology was located at a single site on the mtDNA map from strain M3 (ref. 18; Fig. 1a). Additional cDNAs with homology to this region of the mtDNA were produced and selectively amplified using the polymerase chain reaction (PCR)¹⁹. Sequence analysis of the cDNAs revealed a single reading frame of 528 codons which lacked all three of the classic termination codons. The amino-acid sequence deduced from the open reading frame using the classical genetic code had homology to the α subunit of ATP synthetase from a variety of organisms (Fig. 2).

The sequence of the region of the mtDNA corresponding to the cDNAs was determined. Although the mtDNA and cDNA sequences had a large degree of homology, there was no reading frame in the mtDNA sequence which could code for the ATP synthetase subunit and the regions of reading frame homology were separated by a number of -1 frameshifts. These sequence discrepancies were also reflected in restriction enzyme site differences between the cDNA and the mtDNA.

When the cDNA open reading frame was compared with the analogous sequence of the mtDNA, a single cytidine insertion was found at 54 sites in the cDNA relative to the mtDNA. These insertions generated the open reading frame in the cDNA by correcting the 54 frameshifts on the mtDNA (Fig. 3).

One possible explanation for the discrepancy between the sequence of the cDNA and the mtDNA is that the RNA with the open reading frame was encoded on a conventional DNA template located outside the mitochondrion. Southern blots of DNA from purified *Physarum* nuclei incubated with probes homologous to the cDNA sequence did not detect homologous sequences (data not shown). To screen for a conventional template present at levels below the limits of detection by Southern blotting, PCR was used to amplify DNA sequences with α subunit homology (Fig. 1b). This did not detect any DNA other than contaminating traces of mtDNA which lack the open

TABLE 1 Codons created by RNA editing*

TTT	Phe	22	TCT	Ser	16	TAT	Tyr	23	TGT	Cys	4
TTC	Phe	9 (2)	TCC	Ser	3 (1)	TAC	Tyr	1 (1)	TGC	Cys	1
TTA	Leu	31	TCA	Ser	2	TAA	End	0	TGA	End	0
TGG	Leu	2	TCG	Ser	0	TAG	End	0	TGG	Trp	0
CTT	Leu	13 (1)	CCT	Pro	7	CAT	His	5	CGT	Arg	13
CTC	Leu	2 (1)	CCC	Pro	0	CAC	His	0	CGC	Arg	1
CTA	Leu	5 (2)	CCA	Pro	10 (3)	CAA	Gln	24 (3)	CGA	Arg	2 (1)
CTG	Leu	0	CCG	Pro	2	CAG	Gln	1	CGG	Arg	0
ATT	Ile	31	ACT	Thr	16 (1)	AAT	Asn	23	AGT	Ser	4
ATC	Ile	15 (14)	ACC	Thr	8 (8)	AAC	Asn	2 (1)	AGC	Ser	1
ATA	Ile	1	ACA	Thr	8	AAA	Lys	23	AGA	Arg	7
ATG	Met	8	ACG	Thr	0	AAG	Lys	2	AGG	Arg	1
GTT	Val	15	GCT	Ala	28 (1)	GAT	Asp	34	GGT	Gly	31
GTC	Val	9 (8)	GCC	Ala	8 (5)	GAC	Asp	2 (1)	GGC	Gly	2
GTA	Val	6	GCA	Ala	4	Gaa	Glu	29	GGA	Gly	9
GTG	Val	1	GCG	Ala	0	GAG	Glu	1	GGG	Gly	0

* Cognate amino acids were assigned using the classical genetic code. Numbers directly to the right of the amino acid indicate the number of times the codon is used in the α subunit open reading frame. Numbers in parentheses refer to the number of codons created by C insertion.

reading frame. So, no conventional template with the open reading frame appears to exist in *Physarum* and messenger RNAs for the ATP synthetase α subunit are produced by insertion of cytidine residues in an RNA synthesized on an mtDNA template (RNA editing).

Of the 54 RNA editing sites identified, each is a single residue insertion, each residue is a cytidine, and each is within the reading frame of the α subunit gene. The distribution of the insertion sites is fairly uniform throughout the reading frame of the gene. The average distance between insertion sites is 26 nucleotides with a standard deviation of about 10 nucleotides. The distances range from 11 to 52 nucleotides. Analysis of the relative position of editing sites did not reveal overlapping sets of consistently spaced sites. Comparison of the cDNA and mtDNA sequences flanking the first and last identified editing sites shows that regions of 234 nucleotides (upstream) and 66 nucleotides (downstream) exist which do not have an insertion site, indicating that no insertions occur at the extreme ends of the reading frame nor within the region upstream of the gene. (The region of the cDNA downstream of the termination codons has not been sequenced.)

Although no specific rule for the location of cytidine insertion has been identified, the location of the sites deviates significantly

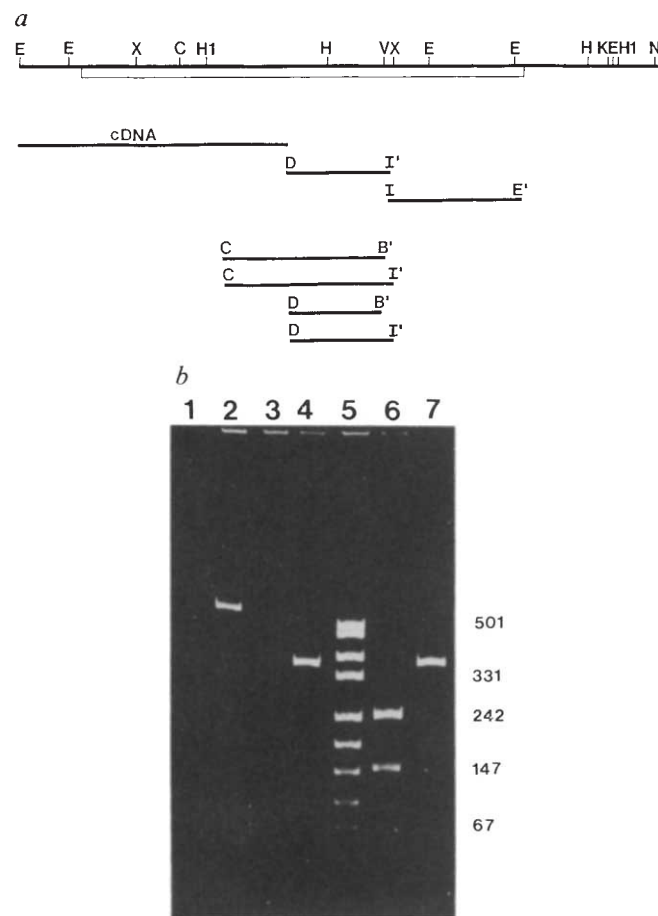
from random. There is a significant preference for sites which follow a purine-pyrimidine dinucleotide. At those sites where the exact position of the cytidine can be identified, 68% follow a purine-pyrimidine dinucleotide. (The exact site of insertion next to a pre-existing cytidine cannot be determined.) All but one of the cytidine insertions in which the exact site of cytidine insertion is ambiguous could follow the purine-pyrimidine rule, and if they do, 78% of the insertions are after purine-pyrimidine dinucleotides.

The insertion sites also show a marked preference for the third position in the codons of the reading frame of the gene (Table 1). Of the insertion sites whose location can be identified exactly, 74% are at the third position. If the remaining sites are assumed to follow purine-pyrimidine dinucleotides, then 78% of the sites are third position. These two biases in the location of the insertion sites result in 88% of the purine-pyrimidine-cytidine codons being created by editing.

The cause of these biases in the location of the editing sites is not known and the biases are not absolute. A number of editing sites are located at positions other than downstream of purine-pyrimidine dinucleotides or at the third position of codons, and numerous locations consistent with these positions are not editing sites.

FIG. 1 Location of PCR amplification products relative to *Physarum* mtDNA. **a**, The top line shows the region of the mtDNA with homology to the cDNAs. The box indicates the region of the mtDNA which has homology to the open reading frame in the cDNAs. Restriction sites are: E, *EcoRI*; X, *XbaI*; C, *Clai*; H1, *HpaI*; H, *HindIII*; V, *PvuII*; K, *KpnI*; and N, *NcoI*. Directly below the map are three cDNAs related to this region of the mtDNA. The cDNA obtained from the λ gt 11 library is labelled 'cDNA'. Two cDNAs labelled DI' and IE' were generated by coupling primer extension to PCR¹⁹ and are designated by the primers used in their synthesis (primers D and I are identical to the sequences in Fig. 3 from 733-752 and 1,100-1,120, respectively; primers I' and E' are complementary to the sequences in Fig. 3 from 1,100-1,120 and 1,555-1,575, respectively). The letters on the right side of the cDNA designate the initial primer used in primer extension. The relative size and location of amplification products produced by PCR using primer sets CB', CI', DB' and DI' (see **b**) are shown below the cDNAs. **b**, PCR amplification products from DNA isolated from *Physarum* nuclei and treated with RNase A. Four sets of primers (CB', CI', DB', DI') were used in PCR amplification¹⁹ (primer C is identical to the sequence in Fig. 3 from 505-525 except that it lacks the cytidine at position 523 added by editing; primer B' is complementary to the sequence in Fig. 3 from 1,063-1,083). Primer B' had homology to a segment of the cDNA which had an extra cytidine relative to the analogous sequence of the mtDNA. Under the conditions used, this primer hybridized with mtDNA very inefficiently. Amplification products made using the four sets of primers were separated on a 6% polyacrylamide gel (lane 1, CB'; lane 2, CI'; lane 3, DB'; and lane 4, DI'). Lane 5 is *HpaI*-digested pUC19. The numbers to the right indicate the size in base pairs (bp) of selected fragments of this digest. PCR amplification of DNA sequences in the nuclear DNA isolation using CI' (lane 2) and DI' (lane 4) produced prominent amplification fragments whereas the sets with the B' primer gave nearly undetectable amounts of the predicted fragments, indicating that the source of the amplified DNA was traces of contaminating mtDNA in the nuclear DNA isolation. To determine the sensitivity of this procedure we intentionally added trace amounts of cloned cDNA to the nuclear DNA preparation. This cDNA was amplified (data not shown) even when the amount of cDNA was 100-fold less (0.05 pg) than that expected for a single-copy gene (about 5 pg). To confirm that the template for the amplified fragments produced by PCR was mtDNA, the 370-bp DI' fragment (lane 4) was digested with the restriction enzymes *HindIII* and *HpaI*. The presence of two cytidine residues in the cDNA sequence creates a *HpaI* site but removes a *HindIII* site relative to the mtDNA sequence. Digestion of the DI' fragment with *HindIII* (lane 6) completely cleaves the 370-bp fragment into two fragments of 235 and 135 bp. Cleavage with *HpaI* (lane 7) would produce fragments of about 255 and 130 bp if the extra cytidine residues were present. No fragments of this size are detected after *HpaI* digestion, confirming that the template for this amplification product was mtDNA.

METHODS. Mitochondrial RNA isolated as described by Jones *et al.*¹⁸ was treated with DNase I (RNase-free, BRL) in the presence of 50 mM sodium acetate, pH 6.5, 10 mM MgCl₂, 2 mM CaCl₂, and 0.5 U Inhibit-ACE (5Prime-3Prime, Inc.) at 37 °C for 30 min. DNase I was removed by phenol extraction. RNA and primer were annealed in 100 mM KCl at 50 °C, precipitated with ethanol, and resuspended in 50 mM Tris-HCl, pH 8.3, 75 mM KCl, 10 mM



dithiothreitol, 3 mM MgCl₂, 50 μ g ml⁻¹ BSA, 25 mM dNTPs and 0.5 U Inhibit-ACE. The primer was then extended using Moloney murine leukaemia virus reverse transcriptase (BRL) at 37 °C for 1 h. cDNAs were treated with RNase A, precipitated with ethanol and amplified using PCR. DNAs for PCR were either cDNAs produced by primer extension or DNA isolated from nuclei purified by the method of Mohberg and Rusch²⁷. DNAs were resuspended in 50 mM KCl, 10 mM Tris-HCl, pH 8.3, 1.5 mM MgCl₂, 0.01% gelatin, 25 μ M dNTPs with 1 μ g of the primers and 2.5 U of *Taq* polymerase (Perkin ElmerCetus). The cycling temperatures and timing were: 94 °C, 1 min, 37 °C, 2 min and 72 °C, 3 min. In the last cycle, extension was carried out for 10 min at 72 °C. Thirty cycles of amplification were performed. cDNAs produced by PCR were cloned in the *HincII* site of pUC18.

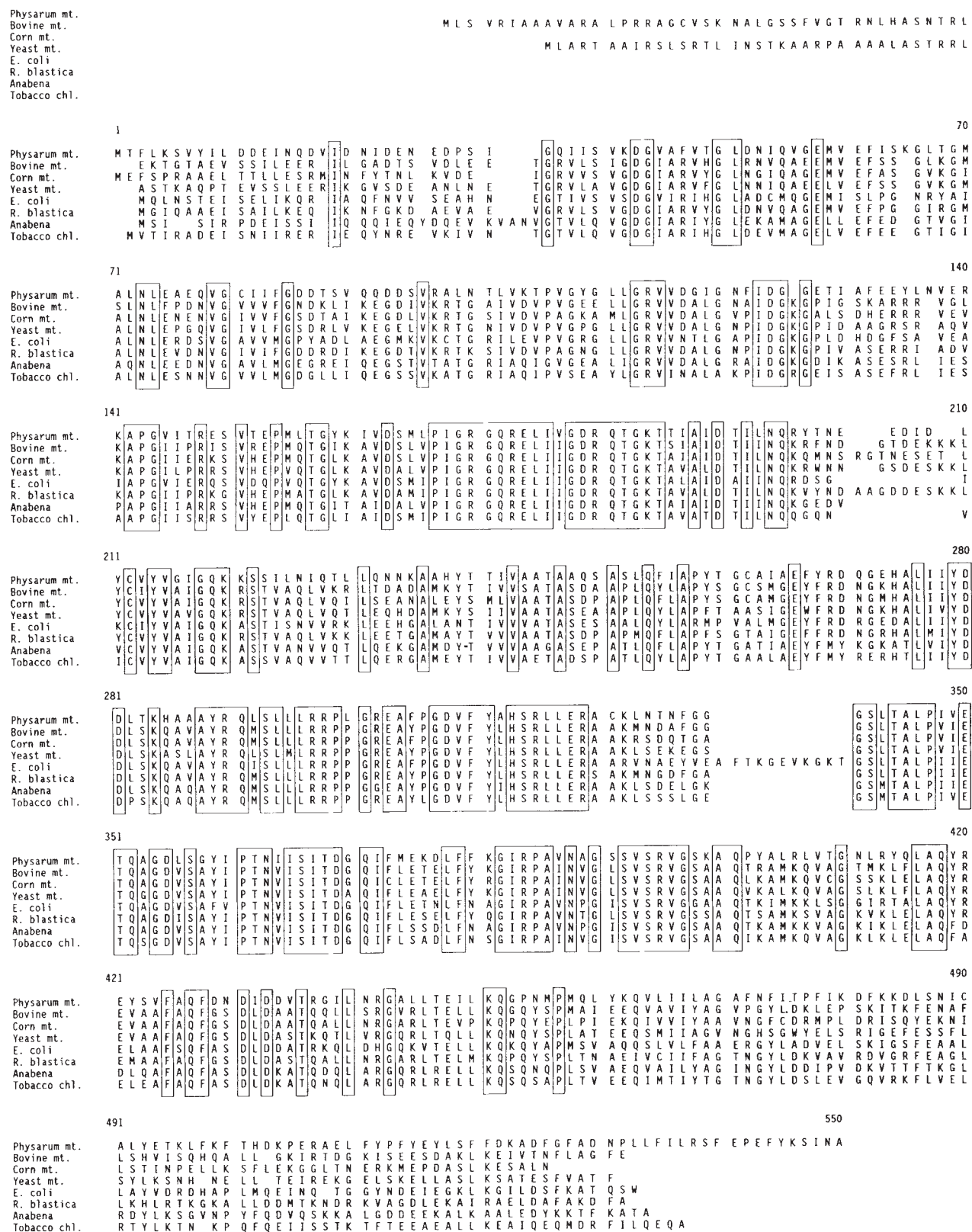


FIG. 2 Alignment of the amino-acid sequence deduced from the open reading frame on *Physarum* mitochondrial cDNAs with the amino-acid sequence of the α -subunit of ATP synthetases from seven different sources. The α subunit sequences compared were from bovine mitochondria²⁸, maize mitochondria²⁹, yeast mitochondria³⁰, *Escherichia coli*³¹, *Rhodospseudomonas blastica*³², *Anabaena* chloroplast³³, and tobacco chloroplast³⁴. Amino-acid sequences are numbered from a methionine in the

Physarum mtDNA reading frame corresponding to the initiation methionine of other α subunits. Amino acids conserved in all eight proteins are boxed. Extension of the amino termini of bovine and yeast mitochondrial α subunits reflect the fact that these mitochondrial proteins are encoded in the nucleus, synthesized on cytoplasmic ribosomes and require this extension for their import into mitochondria.

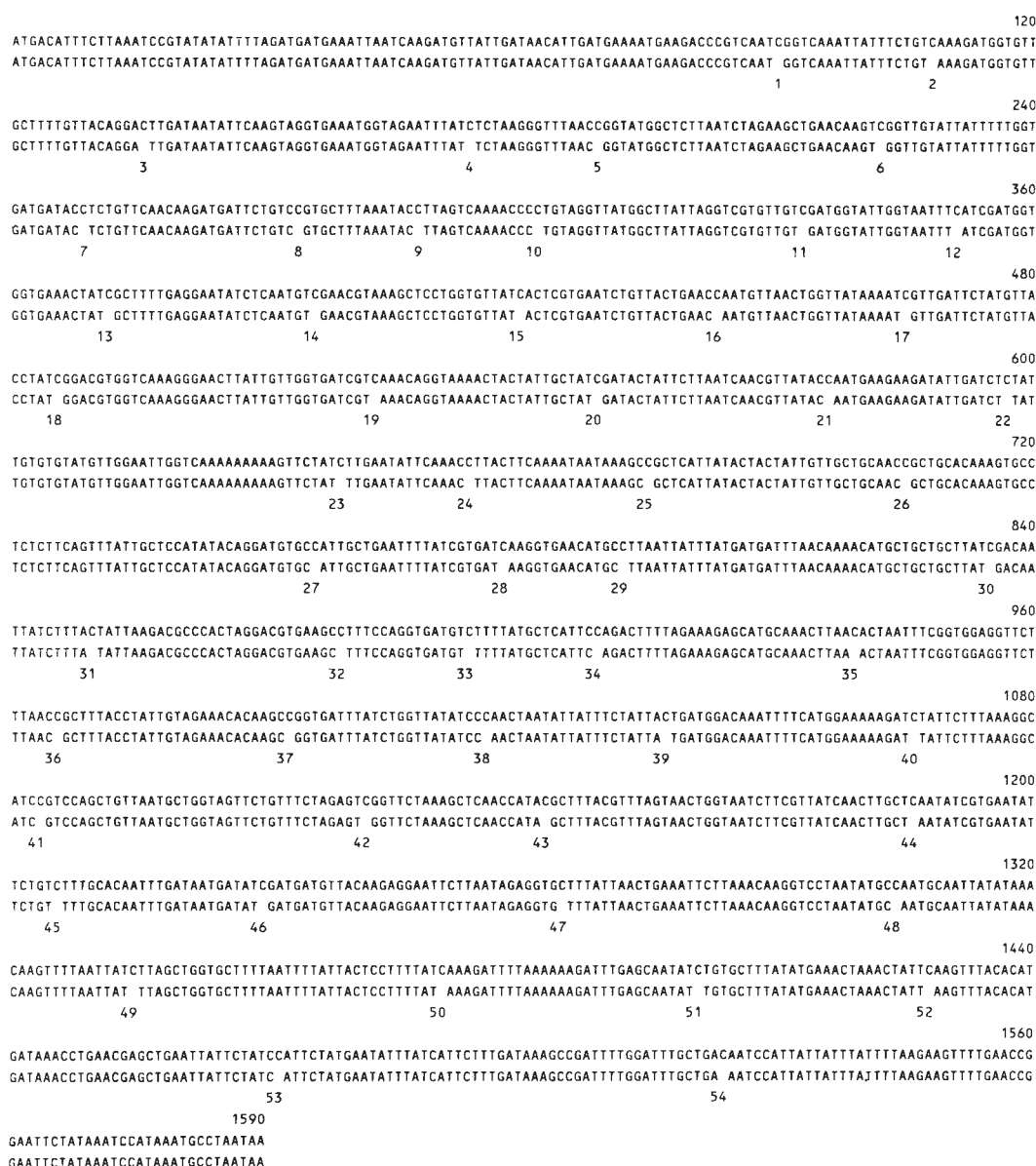


FIG. 3 Alignment of cDNA and mtDNA sequences in the region of the α subunit gene. Sequences are aligned from the ATG initiation codon of the α subunit gene. Sites of cytidine insertion in the cDNA relative to the mtDNA sequence are numbered below the mtDNA sequence. The cDNA residues are numbered from the ATG codon. The GenBank accession numbers for these sequences are M31717 and M31718.

This extensive insertional type of editing is so far unique to the mitochondria of kinetoplastid protozoa and *Physarum*. Substitutional editing has been found in the expression of plant mitochondrial genes⁴⁻⁷, the apolipoprotein B gene⁸⁻¹⁰, measles virus¹¹⁻¹² and in *Xenopus* oocytes¹³⁻¹⁴. This type of editing could be produced by base modification of the type seen during the maturation of transfer RNAs²⁰, by transglycosylation as has been detected for some tRNA modification²¹⁻²⁵, or by exchange (excision/insertion) of a modified or different nucleotide.

Insertional editing has been observed in at least three viral systems but involving only a few residues¹⁵⁻¹⁷. In each of these examples the non-templated residues are inserted within a small homopolymer series (3-4 nucleotides) and this has led to the proposal that the insertion is the result of polymerase slippage or stuttering at specific sites during transcription. Neither base modification nor polymerase stuttering is a viable model for the mechanism of editing observed in the mitochondria of kinetoplastid protozoa or *Physarum*.

The editing systems of kinetoplastid protozoa²⁻⁴ and *Physarum* differ in several ways. Most obvious is the fact that in *Physarum*, cytidines are inserted instead of uridines. Also, the insertions in *Physarum* are fairly evenly distributed within the reading frame and are always a single residue. It remains to be seen whether these two systems achieve RNA

editing by similar mechanisms.

An important question concerning RNA editing is the source of the information specifying the site of insertion. Two possible sources are: (1) the primary sequence or secondary structure of the RNA or (2) a template. The relatively even distribution of editing sites in *Physarum* as well as the fact that only a single cytidine residue is inserted at each site facilitate the search for a consensus sequence or secondary structure in the primary transcript which might specify editing sites. Although the preferential location of editing sites downstream of purine-pyrimidine dinucleotides and at the third position in codons shows that the sites are nonrandom, no consistent rule for the location of the editing sites has been determined. We feel that features of the primary or secondary structure of the transcript are unlikely to specify their location, and we have initiated a search for anti-sense RNAs able to specify editing sites which would be similar to the guide RNAs identified in *Leishmania* by Blum *et al.*²⁶.

The apparent confinement of editing sites to the reading frame of the α subunit is also intriguing. This constraint is lacking in kinetoplastid protozoa as there editing occurs in the poly(A) tail of mRNAs. It will be interesting to see whether genes that encode RNAs lacking reading frames require editing for their expression. Experiments to identify editing in such RNAs are in progress. □

Received 28 September; accepted 9 November 1990.

- Benne, R. *et al. Cell* **46**, 819–826 (1986).
- Benne, R. *Biochim. biophys. Acta* **1007**, 131–139 (1989).
- Simpson, L. & Shaw, J. *Cell* **57**, 355–366 (1989).
- Stuart, K. *Parasitol. Today* **5**, 5–8 (1989).
- Gualberto, J. M., Lammattina, L., Bonnard, G., Weil, J. H. & Grienberger, J. M. *Nature* **341**, 660–662 (1989).
- Covello, P. S. & Gray, M. W. *Nature* **341**, 662–666 (1989).
- Hiesel, R., Wissinger, B., Schuster, W. & Brennicke, A. *Science* **246**, 1632–1634 (1989).
- Chen, S. H. *et al. Science* **238**, 363–366 (1987).
- Hospattankar, A. V., Higuchi, K., Law, S. W., Meglin, N. M. & Brewer, H. B. *Biochim. Biophys. Res. Commun.* **148**, 279–285 (1987).
- Powell, L. M. *et al. Cell* **50**, 831–840 (1987).
- Cattaneo, R. *et al. Cell* **55**, 255–265 (1988).
- Lamb, R. & Dreyfuss, G. *Nature* **337**, 19–20 (1989).
- Bass, B. L. & Weintraub, H. *Cell* **55**, 1089–1098 (1988).
- Kimelman, D. & Kirschner, M. W. *Cell* **59**, 687–696 (1989).
- Schwer, B. & Stunnenberg, H. G. *EMBO J.* **7**, 1183–1190 (1988).
- Thomas, S. M., Lamb, R. A. & Patterson, R. G. *Cell* **54**, 891–902 (1988).
- Cattaneo, R., Kaelin, K., Baczko, K. & Billeter, M. A. *Cell* **56**, 759–764 (1989).
- Jones, E. J., Mahendran, R., Spottswood, M. R., Yang, Y.-C. & Miller, D. L. *Curr. Genet.* **17**, 331–337 (1990).
- Saiki, R. K. *et al. Science* **239**, 487–491 (1988).
- Bjork, G. R. *et al. A. Rev. Biochem.* **56**, 263–287 (1987).
- Elliott, M. S. & Trewyn, R. W. *J. Biol. Chem.* **259**, 2407–2410 (1984).
- Farkas, W. R. & Singh, R. D. *J. Biol. Chem.* **248**, 7780–7785 (1973).
- Howes, N. K. & Farkas, W. R. *J. Biol. Chem.* **253**, 9082–9087 (1978).
- Okada, N. & Nishimura, S. *J. Biol. Chem.* **254**, 3061–3066 (1979).
- Okada, N. *et al. J. Biol. Chem.* **254**, 3067–3073 (1979).
- Blum, B., Bakalara, N. & Simpson, L. *Cell* **60**, 189–198 (1990).
- Mohberg, J. & Rusch, H. P. *Expl Cell Res.* **66**, 305–316 (1971).
- Walker, J. E. *et al. J. molec. Biol.* **184**, 677–701 (1985).
- Braun, C. J. & Levings, C. S. III. *Physiol.* **79**, 571–577 (1985).
- Takeda, M., Chen, W. J., Saltgaber, J. & Douglas, M. J. *Biol. Chem.* **261**, 15126–15133 (1986).
- Walker, J. E., Saraste, M. & Gay, N. J. *Biochim. biophys. Acta* **768**, 164–200 (1984).
- Tybuliewicz, V. L. J., Falk, G. & Walker, J. E. *J. molec. Biol.* **179**, 185–214 (1984).
- McCann, D. F., Whitaker, R. A., Alam, J., Vrba, J. M. & Curtis, S. E. *J. Bact.* **170**, 3448–3458 (1988).
- Deno, H., Shinozaki, K. & Sugiyama, M. *Nucleic Acids Res.* **11**, 2185–2191 (1983).

ACKNOWLEDGEMENTS. We thank V. Vogt (Cornell University) for sharing his *Physarum* cDNA library, and the Texas Advanced Research Program and the National Institute of General Medical Sciences (NIH) for support.

Type I iodothyronine deiodinase is a selenocysteine-containing enzyme

Marla J. Berry, Laila Banu & P. Reed Larsen

Howard Hughes Medical Institute Laboratory and Thyroid Division, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 75 Francis Street, Boston, Massachusetts 02115, USA

ALTHOUGH thyroxine (3,5,3',5'-tetraiodothyronine, T_4) is the principal secretory product of the vertebrate thyroid, its essential metabolic and developmental effects are all mediated by 3,5,3'-triiodothyronine (T_3), which is produced from the prohormone by 5'-deiodination. The type-I iodothyronine deiodinase, a thiol-requiring propylthiouracil-sensitive oxidoreductase, is found mainly in liver and kidney and provides most of the circulating T_3 ¹ but so far this enzyme has not been purified. Using expression cloning in the *Xenopus* oocyte, we have isolated a 2.1-kilobase complementary DNA for this deiodinase from a rat liver cDNA library. The kinetic properties of the protein expressed in transient assay systems, the tissue distribution of the messenger RNA, and its changes with thyroid status, all confirm its identity. We find that the mRNA for this enzyme contains a UGA codon for selenocysteine which is necessary for maximal enzyme activity. This explains why conversion of T_4 to T_3 is impaired in experimental selenium deficiency^{2–6} and identifies an essential role for this trace element in thyroid hormone action.

We and others have reported expression of type-I 5' deiodinase activity in oocytes following injection of a 1.9–2.4-kilobase (kb) fraction of rat liver poly(A)⁺ RNA^{7,8}. We have now constructed a unidirectional, size-fractionated rat liver cDNA library for expression screening in *Xenopus* oocytes. Plasmid DNA was transcribed *in vitro*, the resulting RNA injected into oocytes, and oocyte homogenates assayed for their ability to deiodinate 3,3',5'-triiodothyronine (rT_3). This strategy

TABLE 1 Effect on sensitivity of type-I iodothyronine deiodinase to gold after substitution of cysteine for selenocysteine-126

Transfected cDNA	Gold thioglucose (nM)	Deiodinase activity* of JEG cell sonicates	Inhibition (%)
TGA (selenocysteine-126)	0	7.0	0
	10	4.0	43
	100	0.95	95
TGT (cysteine-126)	0	1.3	0
	10	1.3	0
	100	1.1	13

Wild-type (TGA) or cysteine-mutant (TGT) G21 cDNA in CDM-8 was transfected into JEG-3 cells as described for Fig. 3. Protein from sonicated cells (435 µg, wild type; 680 µg, cysteine mutant) was incubated in 500 µl 0.1M phosphate buffer containing 10 mM DTT, 1 mM EDTA, 0.1 M phosphate buffer and 300 nM [¹²⁵I] rT_3 for 30 min in the presence of the indicated concentrations of gold thioglucose. [¹²⁵I] rT_3 was quantitated as described⁷. 1-β-D-thioglucose (5 µM) had no effect on enzyme activity.

* Deiodinase activity expressed in pmol rT_3 deiodinated per min per mg of protein.

resulted in isolation of a single positive clone, G21. The DNA sequence and predicted amino-acid sequence of the rat liver type-I 5' deiodinase clone G21 are shown in Fig. 1.

To confirm that this clone encoded the type-I 5' deiodinase, it was expressed in JEG-3 human choriocarcinoma cells after DNA transfection. The 2.1-kb insert was excised from Bluescript, inserted into the mammalian expression vector CDM-8⁹, and the resulting construct transfected into JEG-3 cells using the calcium phosphate method¹⁰. After two days we assayed cell homogenates or microsomal fractions for type-I 5' deiodinase activity using rT_3 as substrate⁷. The K_m for rT_3 was 130 nM in the presence of 5 mM dithiothreitol, which agrees with the value for type-I deiodinase in rat liver microsomes¹¹. There was no deiodinase activity in cells transfected with CDM-8 vector alone. Propylthiouracil is a competitive inhibitor of dithiothreitol, giving >50% inhibition at 0.5 µM. T_4 acts as a competitive inhibitor of rT_3 deiodination, and was converted to T_3 by microsomal protein from transfected, but not control, cells.

Clone G21 complementary RNA hybridized to a single band of ~2 kb in mRNA from thyroid, kidney, liver and pituitary, but not in that from spleen, heart, lung, small intestine or brown fat (Fig. 2a). This tissue distribution is in agreement with studies using enzyme assays in tissue homogenates¹¹. Using the *Xenopus* oocyte expression assay system, we have previously shown that alterations in the quantity of deiodinase mRNA in rat liver parallel changes in type-I deiodinase activity and the thyroid status of the animal, being reduced in hypothyroid and increased in hyperthyroid rats⁷. Northern blots confirm that the G21 cRNA-hybridizing 2-kb band also changes with thyroid status (Fig. 2b). Hybridization with a β-actin cRNA probe showed that these differences were not due to generalized alterations in mRNA levels (data not shown).

The DNA sequence of clone G21 predicts a protein of relative molecular mass ~14,000 (14K), initiating at nucleotide 7 and terminating at nucleotide 382. We constructed deletions from the 5' or 3' ends, an internal deletion, and a frameshift insertion to identify regions essential for deiodinase activity. The locations of these mutations and their effects on activity in both oocytes and transfected JEG cells are shown in Fig. 3a. The absence of activity with the *Pst*I 5' deletion confirmed that sequences 5' to nucleotide 56 are necessary for production of active enzyme, indicating that the ATG at position 7 is the initiation codon. The absence of deiodinase activity using the *Acc*I 3' deletion indicates a requirement for sequences beyond nucleotide 427, and thus beyond the putative stop (TGA) codon. The absence of activity with the *Acc*I frameshift construct indicates that sequences beyond nucleotide 427 must be in-frame, and there-

fore are likely to be coding. The *Hind*III internal deletion construct was fully active, indicating that the sequences between nucleotides 746 and 1,360 are not necessary for deiodinase activity.

To confirm the surprising result that sequences beyond the putative stop codon are needed for expression of active deiodinase, the protein products of the type-I deiodinase clone were analysed by *in vitro* transcription followed by translation in wheat germ and rabbit reticulocyte lysates in the presence of [³⁵S]methionine. In the wheat germ system a major labelled protein of ~14K was produced, corresponding to termination at the UGA codon at nucleotide 382 (data not shown). But in the reticulocyte lysate, a ~14K protein and, to a lesser extent, a ~27K protein were both labelled. These results are explained if the UGA codon at nucleotide 382 encodes the rare amino acid selenocysteine, and if the full-length protein terminates at the UAG codon at nucleotide 778. Translation of UGA as selenocysteine has been reported for the mammalian glutathione peroxidase gene^{12,13} and several bacterial genes¹⁴⁻¹⁶. The 27K protein might be missing from wheat germ lysate as a result of

the absence of a selenocysteine transfer RNA_{UGA} in plants¹⁷.

To prove that the UGA codon at position 382 encodes selenocysteine, we converted TGA to the stop codon, TAA, the leucine codon, TTA, or to the cysteine codon, TGT, by oligonucleotide-directed mutagenesis. *In vitro* synthesized mRNAs from these constructs were assayed for expression of type-I 5' deiodinase in oocytes (Fig. 3b) and translated in the presence of [³⁵S]methionine *in vitro* in reticulocyte lysates (Fig. 4). Conversion of the TGA codon to either a stop codon or a leucine codon resulted in complete loss of deiodinase activity, whereas conversion to a cysteine codon resulted in activity of ~10% of the wild-type G21 level. Clone G21 mRNA produced proteins of ~14K and ~27K. The TAA mutation resulted in loss of the ~27K protein, without affecting production of the ~14K protein. The TTA and TGT mutant proteins were ~27K, in close agreement with the molecular weight of 29K predicted by the DNA sequence (Fig. 4). The *Hind*III internal deletion produced a ~26K protein. This construct encodes a protein six amino acids shorter than wild type, deleting the last ten amino acids and substituting four amino acids. The 27K protein is similar in size to a protein identified in rat thyroid, liver and kidney microsomes by labelling with ⁷⁵Se (refs 18, 19). A protein of this size has also been identified by covalent labelling of rat kidney and liver microsomes with [¹²⁵I]bromacetyl T₄ or T₃, which are substrates for the type-I deiodinase^{20,21}.

Our results and the recent demonstration of impaired type-I

```

1  GCTGAGATGGGGCTGTCGCCAGCTATGGCTGTGGCTGAAGCGGCTTGTGATATCTCGCAG
   M G L S Q L W L W L K R L V I F L Q
61  GTAGCGCTGGAGGTGGCTACGGGCAAGGTGCTAATGACACTGTCAGAGAGAGTCAAG
   V A L E V A T G K V L M T L F P P B R V K
121 CAGAACATCTCGCCATGGGCCAAGAACCGGAATGACCGGAATCCCGCATTCGCCOCT
   Q N I L A M G Q K T G M T E N P R F A P
181 GACAACTGGGTCGCCACTCTCTCAGCATCCAGTACTTGTGTGTGCTGAAGTCCGGC
   D N W V P T F F S I Q Y F W F V L K V R
241 TGGCAGAGACTGGAAGACAGGGCTGAGTATGGGGGCTGGCCCCCACTGCCCGTGGTC
   W Q R L E D R A E Y G L A P N C T V G
301 CGCCTCTCAGGACAGAAGTGCACGTCTGGGATTCATTCAAGGCAGCAGACCCCTGGTG
   R L S G Q K C N V W D F I Q G S R P L V
361 TTGAATCTGGCAGCTGCACCTGACCTTCTTCTTCTCAAAATTGACGAGTTCAAGAGA
   L N F G S C T S C P S F L L K F D Q F K R
421 CTCGTAGACGACTTTCCTCCACAGCTGACTTCTCATCATTTACATTGAAGAGCTCAC
   L V D D F A S T A D F L I I Y I E E A H
481 GCCACAGATGGATGGGCTTTTAAAGAACAGCTGGACATCAGGCAGACCGGAGCTCCAG
   A T D G W A F K N N V D I R Q H R S L Q
541 GACCGCTCGGGGAGCAGACATCTGCTGGCCAGGAGCCCGCAGTGTCTGTGGTGGTG
   D R L R A A H L L L A R S P Q C P V V V
601 GACCAATGACAGAACGAGCAGCAGCTCTATGACAGCTGCTGCTGAGAGGCTCTATGTG
   D T M Q N Q S S Q L Y A A L P E R L Y V
661 ATACAGGAAGGAGGATCTGCTACAGGTAACCTGGCCCTTGGAACTACAACCTGAG
   I Q E G R I C Y K G K P G P W N Y N P E
721 GAAGTCCGAGCTGTTCTGAAAAAGCTTTCATCCACCTGGACACATGCTCAGTCTAG
   E V R A V L E K L C I P P G H M P Q F
781 GGGGCCAGCAGGAAGTCCCGCAAGCTTGGTACTCTCCCGCAGCAGTACAGATGCTCTT
   AGCTTTGACCTTCTGTTCCAGATCAATTAAGTCTGAGATTTTCTGATCTGAAACAAATA
901 ACTACCGGGAGGCAATTCAGTTCACAGCAGCCCAACAGCAGCAAAATGTTACAAACAGAG
   ATAAAGCAATACGAGCTGTTAAAGAAAGTAAAGTGTGAGCTTTGACCACTCCACAGG
1021 CGGAGACCAATCCAGTGTGTGCCCCCTTCTGCTGGAAGGGTACTCATGCTTGGTGGCTGA
   CTCTTGAAGTGTAGTGAATCATGATGATGACGCTCAAAAGCTCAATCCATTGGCCCAAGTT
1141 TGCCATCATAGATCAGTCTTGTAGTACCAAGCCAGCAGGCGGCTATTCTACTGTGA
   GCAACCAAGACATTTGAAACACTTTTCTGGCCCTAAGATTGAAATCGGTTAATATTGTT
1261 GGTGATAGGTGTTTCCATGGCAAGCTATAATCTAATCTGCTCCCTCTACCATCTTTGAA
   TAGATTGACAGAAATCTGGCTCTCTGCTGACACAAAAGCTTTATAAGCTTTAACTAA
1381 ACCAAATACAGGCGCCAGCAAAAGCTGCAATCCCGCTGCTGTAACCTGTGTCACCTGGC
   GCCAGTCTCTTACTGCTCTTTCATGTTAGATGGCTTTGACCTGACGGGTAGCCATGGGT
1501 TCATCTGTCTGCTGCTCTCTTTTATATTTTATGTTATGATGGTCAAGTGTAAAGTTTCA
   CACGCTGTGACTTGTATTTTAAAAATGTGGGAAAGTGCAGCAAGCTAACGATTTAAATC
1621 CCGCAGGCTATTTTGAATGGCTCCGGTGTGATCCTTACAAATTCCTTTCTGAGTTTGT
   ATGTGGGCTGCTCTGCTGCTTTTCCGATAGCCACGCTGTAATGTAATCAGCTAAGGCA
1741 TCGTTTGGCTGGAGGAGCCCGCTGAGGAGAAAGCTGCTATGTTGGCAGCAGTCACAC
   ATGTTGCTCTGGAAGTGTGGAAGGAGCTGGCTGTTACGCTCAGCAAGCAAGCACT
1861 TTAGGGGTGATGCTGAATGGACCTGGGAGCATTCTCCAGGCATCCAAACAGTTCTCTC
   TTGCTCTGCTTAGGCTACACCAATACTGTAACATTCGATTATGATGAGATTAGGT
1981 GAGTCAGATCTAGCTATAAGTGCAGAGTGGCTGTAACCTTCAAACTCTCAGACTCAGA
   GTAGCTGGGATTCCAGCTCTGCCCCCTATAAAAAATGCTTTGACCTCTTGAAAAAA
2101 AAAAAA

```

FIG. 1 DNA sequence and predicted amino-acid sequence of type-I 5' deiodinase.

METHODS. cDNA synthesis was catalysed by avian myeloblastosis virus (AMV) reverse transcriptase (Life Sciences)³⁰. Double-stranded cDNA was size-fractionated on low-melting-temperature agarose (Sea Plaque, FMC) and the region corresponding to 1.8–2.5 kb isolated. The resulting cDNA was ligated to adaptors (*In Vitrogen*), inserted into lambda Zap II (Stratagene), and packaged *in vitro*. The library was subdivided, amplified, and converted to Bluescript plasmid by *in vivo* excision as described in the procedures of Stratagene. Plasmid DNA was linearized and transcribed *in vitro* using T7 RNA polymerase. *Xenopus laevis* oocytes were manually dissected, injected with *in vitro*-transcribed RNA (0.5 to 20 ng per oocyte in 40 nl diethylpyrocabonate H₂O), and incubated for 3 days at 18°C in 50% Leibovitz's L-15 medium, 15 mM HEPES, 100 µg ml⁻¹ gentamycin and 50 units ml⁻¹ nystatin. Oocyte homogenates were assayed for type-I 5' deiodinase as described⁷. Upper and lower DNA strands were sequenced by the dideoxy method³¹ using a T7 sequencing kit from Pharmacia.

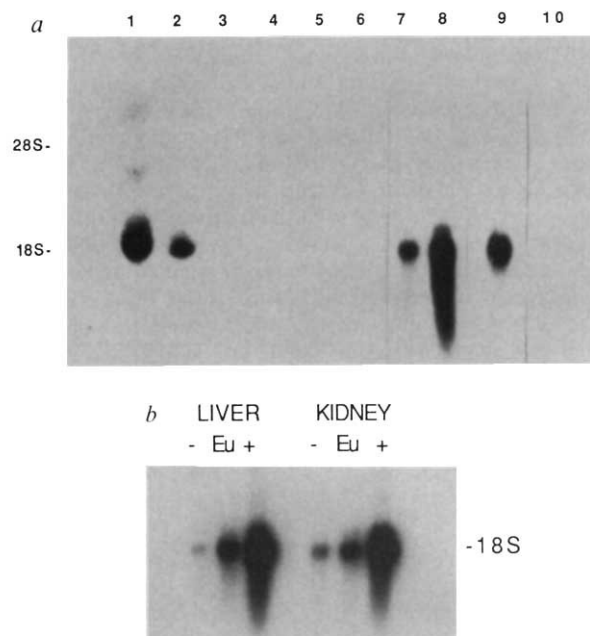


FIG. 2 Northern blot analysis of type-I 5' deiodinase mRNA. *a*, Tissue distribution. RNA was isolated from rat tissues by standard guanidinium thiocyanate methods⁷. Total RNA or poly(A)⁺ RNA was electrophoresed on 1.1% agarose formaldehyde gels. Blots were probed with clone G21 cRNA or β -actin cRNA and washed at high stringency. Lanes 1–6 contain 20 µg total RNA from kidney, liver, spleen, heart, lung and small intestine, respectively. Hybridization to a β -actin probe showed that there was significantly less mRNA in liver than in the other samples, accounting for the lower signal in this lane. Lanes 7 and 8 contain 2 µg poly(A)⁺ RNA, from the thyroids of methimazole-treated (hypothyroid) rats (lane 7) and from kidney (lane 8). Lanes 9 and 10 contain 5 µg poly(A)⁺ RNA from pituitary and brown adipose tissue, respectively. Autoradiography of blots of lanes 1–6 was for 4 days, lanes 7 and 8 for 1 hour, and lanes 9 and 10 for 1 week. *b*, Effect of thyroid status on type-I 5' deiodinase mRNA levels. Rats were made hypothyroid by treatment for 3 weeks with 0.02% methimazole in the drinking water. Hyperthyroidism was induced by intraperitoneal injection of 50 µg T₃ daily for three days. Liver and kidney poly(A)⁺ RNA (5 µg) from hypothyroid ('minus' lanes), euthyroid (Eu lanes), and hyperthyroid ('plus' lanes) rats was probed with G21 cRNA. Blots were autoradiographed for 1 h. Migration positions of ribosomal 18S and 28S RNA markers are shown.

FIG. 3 Expression of type-I 5' deiodinase from G21 wild-type and mutant constructs. *a*, Deletions were constructed by digestion with restriction enzymes at the indicated sites in clone G21 and convenient sites in the vectors, followed by purification on agarose gels of the fragments, religation, and mapping of the resulting constructs. The *AccI* frameshift was constructed by digestion with *AccI*, followed by conversion to blunt ends with DNA polymerase large fragment, and religation. All mutations were confirmed by DNA sequencing. RNAs were transcribed *in vitro* and injected at 0.1 to 20 ng per oocyte. DNA transfections¹⁰ and deiodinase assays⁷ have been described. Assay of human growth hormone in medium from a cotransfected human growth hormone-expressing plasmid confirmed that transfection efficiencies were equivalent¹⁰. Oocyte activity of 100% is defined as deiodination of 30–40% of 2 nM [¹²⁵I]rT₃ per hour with a homogenate of 4 oocytes injected with 0.1 ng G21 RNA per oocyte. G21 RNA was at least 100-fold more active per ng than liver poly(A)⁺ RNA. JEG-3 cells were homogenized and 100 to 300 µg protein from homogenates or 20,000 pellets were incubated in 400 µl containing 25 mM DTT and 5 nM [¹²⁵I]rT₃ for 1 hour. [¹²⁵I][–] was quantitated as described⁷. Equal quantities of I[–] and 3,3'-diiodothyronine are produced during this reaction. All assays were in duplicate. ND, not determined. *b*, The TGA codon at nucleotide 382 was replaced by the codons indicated using the P-select *in vitro* mutagenesis

deiodinase activity in selenium-deficient rats^{2,4,5} implicate this trace element as the nucleophilic atom in the active site of the type-I iodothyronine deiodinase. This is analogous to the selenocysteine in the active site of mammalian glutathione peroxidase²², the only previously identified eukaryotic selenocysteine-containing enzyme. From Table 1 it can be seen that the type-I deiodinase shares another property with glutathione peroxidase, namely sensitivity to inhibition by gold, which is thought to complex with the selenolate group in the active site of this enzyme²³. The activity of the transiently expressed wild-type deiodinase protein is inhibited by ~50% by 10 nM gold thioglucose. Substitution of cysteine for selenocysteine resulted in an enzyme with ~20% of the intrinsic activity of the wild-type protein, in agreement with the oocyte studies. This mutant protein was much less sensitive to inhibition by gold thioglucose than the native enzyme. In experiments not shown, ~50% inhibition of the cysteine-substituted enzyme was observed at 1.0 µM gold thioglucose.

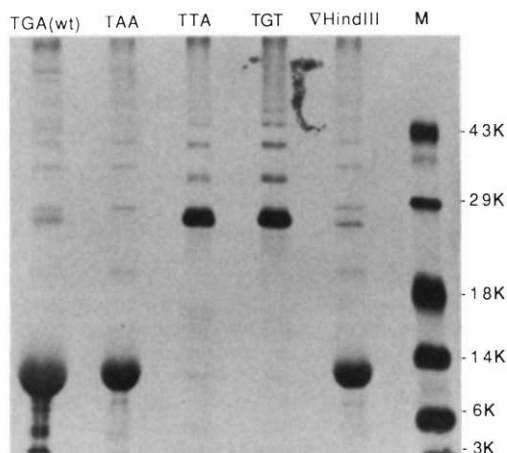
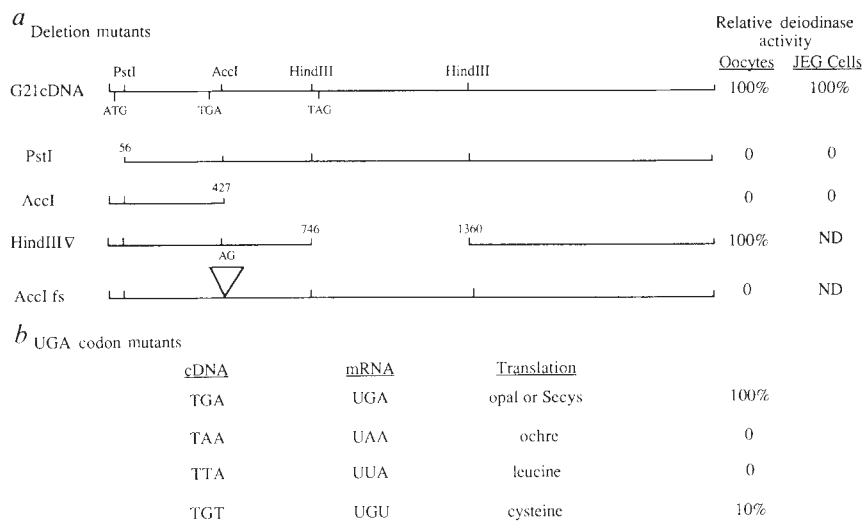


FIG. 4 *In vitro* translation and SDS-polyacrylamide gel electrophoresis of type-I deiodinase constructs. Construction and *in vitro* transcription of G21, substitution mutants, and the HindIII internal deletion are described in the legends to Figs 1 and 3. *In vitro*-transcribed RNA was translated in rabbit reticulocyte lysates (Promega) using [³⁵S]methionine according to the manufacturer's protocol. *In vitro* translation products were analysed on a 15% polyacrylamide gel and visualized by fluorography after impregnation of the gel with EN³HANCE (DuPont). Sizes of molecular weight markers (lane M) are indicated on the right.



system of Promega. Briefly, the insert from clone G21 was ligated into P-select and single-stranded phagemid DNA was obtained. Oligonucleotides corresponding to the changes were annealed to the single-stranded DNA and double-stranded DNA synthesized. The entire coding regions of these plasmids were sequenced to confirm that these were the only mutations.

Type-I deiodinase and glutathione peroxidase both exhibit ping-pong kinetics with reduced thiols as cosubstrate, and can be inhibited by carboxymethylation^{11,22,24,25}. However, we found no significant homology between the two enzymes, nor to any other protein sequence in GenBank or EMBL²⁶. This includes protein disulphide isomerase, another thiol-requiring protein which is possibly related to the type-I deiodinase²⁷. This result is consistent with our earlier expression studies and with the biochemistry of the two enzymes^{7,21}. Thyroxine must be converted to T₃ to be active *in vivo*, and it is the type-I iodothyronine deiodinase which is responsible for most of this conversion^{28,29}. Our results explain why selenium is essential in thyroid hormone action. □

Received 28 September; accepted 5 December 1990.

- Larsen, P. R., Silva, J. E. & Kaplan, M. M. *Endocr. Rev.* **2**, 87–102 (1981).
- Beckett, G. J., Beddows, S. E., Morrice, P. C., Nicol, F. & Arthur, J. R. *Biochem. J.* **248**, 443–447 (1987).
- Arthur, J. R., Morrice, P. C. & Beckett, G. J. *Res. Vet. Sci.* **45**, 122–123 (1988).
- Beckett, G. J., MacDougall, D. A., Nicol, F. & Arthur, J. R. *Biochem. J.* **259**, 887–892 (1989).
- Arthur, J. R., Nicol, F., Hutchinson, A. R. & Beckett, G. J. *J. inorg. Biochem.* **39**, 101–108 (1990).
- Arthur, J. R., Nicol, F., Rae, P. W. H. & Beckett, G. J. *Clin. Chem. Enzym. Commun.* **3**, 209–214 (1990).
- Berry, M. J., Kates, A.-L. & Larsen, P. R. *Molec. Endocr.* **4**, 743–748 (1990).
- St Germain, D. L. & Morganelli, C. M. *J. Biol. Chem.* **264**, 3054–3056 (1989).
- Aruffo, A. & Seed, B. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8573–8577 (1987).
- Brent, G. A. *et al. Molec. Endocr.* **3**, 1996–2004.
- Leonard, J. L. & Visser, T. J. in *Thyroid Hormone Metabolism* (ed. Hennemann, G.) 189–229 (Dekker, New York, 1986).
- Chambers, I. *et al. EMBO J.* **5**, 1221–1227 (1986).
- Mullenbach, G. T. *et al. Protein Engng.* **2**, 239–246 (1988).
- Zinoni, F., Birkmann, A., Stadtman, T. C. & Bock, A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4650–4654 (1986).
- Menon, N. K., Peck, H. D. Jr, LeGall, J. & Przybyla, A. E. *J. Bact.* **169**, 5401–5407 (1987).
- Ishida, K., Morino, T., Takagi, K. & Sukenaga, Y. *Nucleic Acids Res.* **15**, 10051 (1987).
- Lee, B. J. *et al. Molec. cell. Biol.* **10**, 1940–1949 (1990).
- Behne, D., Hilmert, H., Scheid, S., Gessner, H. & Elger, W. *Biochim. biophys. Acta* **966**, 12–21 (1988).
- Behne, D., Scheid, S., Kyriakopoulos, A. & Hilmert, H. *Biochim. biophys. Acta* **1033**, 219–225 (1990).
- Kohrle, J., Rasmussen, U., Rokos, H., Leonard, J. L. & Hesch, R. D. *J. Biol. Chem.* **265**, 6146–6154 (1990).
- Schoenmakers, C. H. H., Pigman, I. G. A. J., Hawkins, H. C., Freedman, R. B. & Visser, T. J. *Biochem. biophys. Res. Commun.* **162**, 857–866 (1989).
- Forstrom, J. W., Zakowski, J. J. & Tappel, A. L. *Biochemistry* **17**, 2639–2644 (1978).
- Chaudiere, J. & Tappel, A. L. *J. inorg. Biochem.* **20**, 313–325 (1984).
- Tappel, A. L. *Curr. Top. Cell Reg.* **24**, 87–97 (1984).
- Visser, T. J., Frank, S. & Leonard, J. L. *Molec. cell. Endocr.* **33**, 321–327 (1983).
- Devereux, J., Haerberli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387–395 (1984).
- Boada, R. J., Campbell, D. A. & Chopra, I. J. *Biochem. biophys. Res. Commun.* **155**, 1297–1304 (1988).
- Oppenheimer, J. H., Schwartz, H. L. & Surks, M. I. *J. clin. Invest.* **51**, 2493–2497 (1972).
- Frumess, R. D. & Larsen, P. R. *Metabolism* **24**, 547–554 (1975).
- Gubler, U. & Hoffman, B. *Gene* **25**, 263–269 (1983).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).

ACKNOWLEDGEMENTS. We thank B. Seed, D. Moore and M. Hediger for discussion; Y. Chen, D. Kieffer, S. Mandel, N. Moolten and R. Warne for assistance; and M. Flemming for oligonucleotides. This study was supported in part by a grant from the NIH.